GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:27:20 ; Search time 8.32 Seconds
(without alignments)
150.299 Million cell updates/sec

Title: US-09-786-214A-14
Perfect score: 65
Sequence: 1 LPAVVGLSPGEQE 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	67.7	381	2	F75270	cytochrome P450 -
2	43	66.2	821	2	C84304	DNA helicase [impo
3	41	63.1	156	2	F87551	conserved hypothet
4	41	63.1	326	2	T45226	probable N5,N10-me
5	40	61.5	243	2	I54459	MHC H-2K1-k - mous
6	40	61.5	428	2	AG1304	uracil permease ho
7	40	61.5	428	2	AG1676	uracil permease ho
8	40	61.5	540	2	A75250	carboxylesterase,
9	40	61.5	623	2	T40991	probable lysophosp
10	39	60.0	43	2	S21065	Ig kappa chain V r
11	39	60.0	96	2	S45441	Ig kappa chain V r
12	39	60.0	103	2	S19975	Ig kappa chain V r
13	39	60.0	106	2	PS0070	Ig kappa chain V r
14	39	60.0	106	2	PC4282	Ig kappa chain (an
15	39	60.0	107	2	S57444	Ig kappa chain V-J
16	39	60.0	108	2	C30502	Ig kappa chain V r
17	39	60.0	108	2	S33988	Ig kappa chain V r
18	39	60.0	108	2	G44151	Ig kappa chain V r
19	39	60.0	111	2	S23628	Ig kappa chain V r
20	39	60.0	114	2	S54905	Ig kappa chain V r
21	39	60.0	115	1	K3HUVG	Ig kappa chain pre
22	39	60.0	115	1	KVMSL7	Ig kappa chain pre
23	39	60.0	115	2	S11697	Ig kappa chain pre
24	39	60.0	116	2	B25521	Ig kappa chain pre
25	39	60.0	119	2	S41816	Ig kappa chain V r
26	39	60.0	125	2	S40344	Ig kappa chain V-J
27	39	60.0	128	2	PN0445	Ig kappa chain pre
28	39	60.0	128	2	S40379	Ig kappa chain V-J
29	39	60.0	128	2	A56701	Ig kappa chain V r
30	39	60.0	129	2	S29627	Ig kappa chain V r
31	39	60.0	129	2	S40363	Ig kappa chain - h
32	39	60.0	132	2	S05268	Ig kappa chain pre
33	39	60.0	144	2	PL0106	Ig kappa chain pre
34	39	60.0	144	2	B30502	Ig heavy chain V r
35	39	60.0	215	2	A23746	Ig kappa chain V-I
36	39	60.0	319	2	AD0941	probable regulator
37	39	60.0	508	2	E70764	probable cobI prot
38	39	60.0	563	2	JQ0623	nerve growth facto
39	39	60.0	1367	2	S74285	BUD3 protein - yea
40	39	60.0	1506	2	JC5985	phosphoinositide 3
41	38.5	59.2	656	2	E75468	hypothetical prote
42	38	58.5	86	2	S20649	Ig heavy chain V r
43	38	58.5	110	2	S60591	Ig light chain var
44	38	58.5	111	2	PN0537	Ig kappa chain V r
45	38	58.5	152	2	G75184	probable transcrip

ALIGNMENTS

RESULT 1
F75270
cytochrome P450 - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: F75270
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; , M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M. S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: F75270
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-381 <WHI>
A;Cross-references: GB:AE002076; GB:AE000513; NID:g6460285; PIDN:AAF12016.1; PID:g64602
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2473
A;Map position: 1

Query Match 67.7%; Score 44; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPAVVGLSP 9
|||||||||
Db 52 LPAVVGLSP 60

RESULT 2
C84304
DNA helicase [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: C84304
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, ; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabl Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; L
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: C84304
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-821 <STO>
A;Cross-references: GB:AE004437; NID:g10580995; PIDN:AAG19799.1; GSPDB:GN00138
C;Genetics:
A;Gene: hel

```
  Query Match             66.2%;  Score 43;  DB 2;  Length 821;
  Best Local Similarity   90.0%;  Pred. No. 28;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AVVGLSPGEQ 12
              ||||||| ||
Db        326 AVVGLSPAEQ 335


RESULT 3
F87551
conserved hypothetical protein CC2439 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: F87551
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: F87551
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-156 <STO>
A;Cross-references: GB:AE005673; NID:g13423984; PIDN:AAK24410.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC2439
C;Superfamily: Haemophilus influenzae conserved hypothetical protein HI0305

  Query Match             63.1%;  Score 41;  DB 2;  Length 156;
  Best Local Similarity   80.0%;  Pred. No. 12;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AVVGLSPGEQ 12
              ||||| |||:
Db         18 AVVGLDPGEK 27


RESULT 4
T45226
probable N5,N10-methylene-tetrahydromethanopterin reductase (F420-dependent) [imported]
C;Species: Methanolobus tindarius
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C;Accession: T45226
R;Westenberg, D.J.; Braune, A.; Ruppert, C.; Mueller, V.; Herzberg, C.; Gottschalk, G.;
submitted to the EMBL Data Library, September 1998
A;Description: The F420H2-dehydrogenase from Methanolobus tindarius: Cloning of the ffd
A;Reference number: Z22947
A;Accession: T45226
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-326 <WES>
A;Cross-references: EMBL:AJ011519; PIDN:CAB56639.1
A;Experimental source: DSM 2278
C;Genetics:
A;Gene: ffdA

  Query Match             63.1%;  Score 41;  DB 2;  Length 326;
  Best Local Similarity   70.0%;  Pred. No. 24;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AVVGLSPGEQ 12
              |::|| ||||
Db         88 AILGLGPGEQ 97


RESULT 5
I54459
MHC H-2K1-k - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C;Accession: I54459
R;Watts, S.; Davis, A.C.; Goodenow, R.S.
Immunogenetics 29, 355-357, 1989
A;Title: Sequence analysis of the C3H H-2K1-k gene: Relationship to the H-2 loci.
A;Reference number: I54459; MUID:89233303; PMID:2714856
A;Accession: I54459
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-243 <RES>
A;Cross-references: GB:M27134; NID:g199435; PIDN:AAA39610.1; PID:g387456
C;Genetics:
A;Introns: 22/1; 112/1

  Query Match             61.5%;  Score 40;  DB 2;  Length 243;
  Best Local Similarity   77.8%;  Pred. No. 27;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 VGLSPGEQE 13
              :||||||:|
Db        216 LGLSPGEEE 224


RESULT 6
AG1304
uracil permease homolog pyrP [imported] - Listeria monocytogenes (strain EGD-e)
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AG1304
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; M
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AG1304
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-428 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC99917.1; PID:g16411293; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: pyrP
C;Superfamily: uracil transport protein uraA

  Query Match             61.5%;  Score 40;  DB 2;  Length 428;
  Best Local Similarity   70.0%;  Pred. No. 47;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPG 10
              :|:| |||||
Db         42 VPSVTGLSPG 51


RESULT 7
AG1676
uracil permease homolog pyrP [imported] - Listeria innocua (strain Clip11262)
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AG1676
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; M
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AG1676
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-428 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC97183.1; PID:g16414454; GSPDB:GN00178
```

A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: pyrP
C;Superfamily: uracil transport protein uraA

Query Match 61.5%; Score 40; DB 2; Length 428;
Best Local Similarity 70.0%; Pred. No. 47;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPG 10
:|:| |||||
Db 42 VPSVTGLSPG 51

RESULT 8
A75250
carboxylesterase, type B - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: A75250
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; , M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: A75250
A;Molecule type: DNA
A;Residues: 1-540 <WHI>
A;Cross-references: GB:AE002092; GB:AE000513; NID:g6460455; PIDN:AAF12163.1; PID:g646045
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2626
A;Map position: 1
C;Superfamily: cholinesterase; cholinesterase homology

Query Match 61.5%; Score 40; DB 2; Length 540;
Best Local Similarity 70.0%; Pred. No. 60;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGE 11
| |:||:|||
Db 512 PQVLGLAPGE 521

RESULT 9
T40991
probable lysophospholipase precursor - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C;Accession: T40991
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, March 1999
A;Reference number: Z21962
A;Accession: T40991
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-623 <LYN>
A;Cross-references: EMBL:AL049559; PIDN:CAB40176.1; GSPDB:GN00068; SPDB:SPCC1450.09c
A;Experimental source: strain 972h-; cosmid c1450
C;Genetics:
A;Gene: SPDB:SPCC1450.09c
A;Map position: 3
C;Superfamily: yeast lysophospholipase

Query Match 61.5%; Score 40; DB 2; Length 623;
Best Local Similarity 75.0%; Pred. No. 69;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQE 13
|| ||| ||||
Db 76 PASDGLSTGEQE 87

RESULT 10
S21065
Ig kappa chain V region (anti-RH(D)) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 22-Nov-1993 #sequence_revision 24-May-1996 #text_change 09-May-1997
C;Accession: S21065
R;Dlouha, A.; Lecroisey, A.; Henschen, A.; Rouger, P.; Keil, B.
Protein Seq. Data Anal. 4, 317-318, 1991
A;Title: Subgroup assignement of a human monoclonal anti-Rh(D) antibody.
A;Reference number: S21065; MUID:92253544; PMID:1812483
A;Accession: S21065
A;Molecule type: protein
A;Residues: 1-43 <DLO>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin

Query Match 60.0%; Score 39; DB 2; Length 43;
Best Local Similarity 63.6%; Pred. No. 6.9;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQ 12
|| : |||||:
Db 8 PATLSLSPGER 18

RESULT 11
S45441
Ig kappa chain V region - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 21-Jan-2000
C;Accession: S45441
R;Cox, J.P.L.; Tomlinson, I.M.; Winter, G.
Eur. J. Immunol. 24, 827-836, 1994
A;Title: A directory of human germ-line V(kappa) segments reveals a strong bias in thei
A;Reference number: S45324; MUID:94200218; PMID:8149953
A;Accession: S45441
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-96 <COX>
A;Cross-references: EMBL:Z27500
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin
F;16-91/Domain: immunoglobulin homology <IMM>

Query Match 60.0%; Score 39; DB 2; Length 96;
Best Local Similarity 63.6%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQ 12
|| : |||||:
Db 8 PATLSLSPGER 18

RESULT 12
S19975
Ig kappa chain V region (M-T408) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 21-Jan-2000
C;Accession: S19975
R;Weissenhorn, W.; Riethmueller, G.; Weiss, E.M.; Rieber, E.P.
submitted to the EMBL Data Library, March 1992
A;Description: Structural characterization of CD4 mAb.
A;Reference number: S19963
A;Accession: S19975
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-103 <WEI>
A;Cross-references: EMBL:X65097; NID:g52296; PIDN:CAA46225.1; PID:g52297
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin
F;11-85/Domain: immunoglobulin homology <IMM>

Query Match 60.0%; Score 39; DB 2; Length 103;
Best Local Similarity 63.6%; Pred. No. 17;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
Qy          2 PAVVGLSPGEQ 12
              || : |||||:
Db          3 PATLSLSPGER 13
```

RESULT 13
PS0070
Ig kappa chain V region (38C13.V6.1) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 21-Jan-2000
C;Accession: PS0070
R;Levy, S.; Campbell, M.J.; Levy, R.
J. Exp. Med. 170, 1-13, 1989
A;Title: Functional immunoglobulin light chain genes are replaced by ongoing rearrangeme
A;Reference number: A92781; MUID:89310348; PMID:2501443
A;Accession: PS0070
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-106 <LEV>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin
F;16-89/Domain: immunoglobulin homology <IMM>

Query Match 60.0%; Score 39; DB 2; Length 106;
Best Local Similarity 54.5%; Pred. No. 17;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```
Qy          2 PAVVGLSPGEQ 12
              ||::  ||||:
Db          8 PAIIAASPGEK 18
```

RESULT 14
PC4282
Ig kappa chain (anti-SS-A/Ro 60K peptide) (E-42 and E-56) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 07-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 21-Jan-2000
C;Accession: PC4282; PC4284
R;Suzuki, H.; Takemura, H.; Suzuki, M.; Sekine, Y.; Kashiwagi, H.
Biochem. Biophys. Res. Commun. 232, 101-106, 1997
A;Title: Molecular cloning of anti-ss-A/Ro 60-kDa peptide fab fragments from infiltratin
A;Reference number: PC4279; MUID:97236289; PMID:9125110
A;Accession: PC4282
A;Molecule type: protein
A;Residues: 1-106 <SUZ>
A;Note: E-42
A;Accession: PC4284
A;Molecule type: protein
A;Residues: 1-106 <SU2>
A;Note: E-56
C;Comment: This antibody is commonly found in systemic autoimmune diseases such as Sjoeg
C;Superfamily: immunoglobulin V region; immunoglobulin homology
F;14-88/Domain: immunoglobulin homology <IMM>

Query Match 60.0%; Score 39; DB 2; Length 106;
Best Local Similarity 63.6%; Pred. No. 17;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
Qy          2 PAVVGLSPGEQ 12
              || : |||||:
Db          6 PATLSLSPGER 16
```

RESULT 15
S57444
Ig kappa chain V-J region - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 10-Oct-1995 #sequence_revision 17-Nov-1995 #text_change 21-Jan-2000
C;Accession: S57444
R;Paterson, G.; Wilson, G.; Kennedy, P.G.E.; Willison, H.J.
submitted to the EMBL Data Library, June 1995
A;Description: Analysis of anti-GM1 ganglioside IgM antibodies cloned from motor neurop
A;Reference number: S57408
A;Accession: S57444
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-107 <PAT>
A;Cross-references: EMBL:X87898; NID:g871275; PIDN:CAA61149.1; PID:g871276
C;Superfamily: immunoglobulin V region; immunoglobulin homology
F;16-90/Domain: immunoglobulin homology <IMM>

Query Match 60.0%; Score 39; DB 2; Length 107;
Best Local Similarity 63.6%; Pred. No. 17;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
Qy          2 PAVVGLSPGEQ 12
              || : |||||:
Db          8 PATLSLSPGER 18
```

Search completed: May 7, 2004, 12:39:07
Job time : 8.48667 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:20:30 ; Search time 4.68 Seconds
(without alignments)
144.639 Million cell updates/sec

Title: US-09-786-214A-13
Perfect score: 68
Sequence: 1 PAVVGLSPGEQEY 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	63.2	211	1	COBL_METJA	Q58917 methanococc
2	43	63.2	633	1	PLB5_SCHPO	Q9y7n6 schizosacch
3	41	60.3	156	1	RUVX_CAUCR	Q9a5k8 caulobacter
4	41	60.3	326	1	MER_METTI	Q9uxp0 methanolobu
5	41	60.3	508	1	COBI_MYCTU	Q10677 mycobacteri
6	40	58.8	673	1	PLB4_SCHPO	Q9p327 schizosacch
7	39	57.4	115	1	KV3I_HUMAN	P04433 homo sapien
8	39	57.4	115	1	KV5I_MOUSE	P01642 mus musculu
9	39	57.4	1636	1	BUD3_YEAST	P25558 saccharomyc
10	38	55.9	429	1	RS1_LEULA	P50889 leuconostoc
11	38	55.9	507	1	CATA_PICAN	P30263 pichia angu
12	37	54.4	446	1	COBJ_ARCFU	O29534 a cobalamin
13	37	54.4	446	1	ENO1_MAIZE	P26301 zea mays (m
14	37	54.4	597	1	NR41_RAT	P22829 rattus norv
15	37	54.4	607	1	GLMS_CLOTE	Q890u2 c glucosami
16	37	54.4	813	1	CADM_MOUSE	Q9wtp5 mus musculu
17	37	54.4	1402	1	N160_MOUSE	Q9z0w3 mus musculu
18	36.5	53.7	1121	1	BMS1_SCHPO	O94653 schizosacch
19	36	52.9	129	1	KV3H_HUMAN	P04207 homo sapien
20	36	52.9	158	1	RSD_ECOLI	P31690 escherichia
21	36	52.9	162	1	RSD_SALTY	Q9l9i6 salmonella
22	36	52.9	243	1	SUMT_SYNP7	P42451 synechococc
23	36	52.9	279	1	TYSY_CAUCR	Q9a6h0 caulobacter
24	36	52.9	317	1	OAD1_HUMAN	Q8nge0 homo sapien
25	36	52.9	342	1	HUPK_AZOVI	P31878 azotobacter
26	36	52.9	348	1	HOXV_AZOVI	P40597 azotobacter
27	36	52.9	422	1	TKSU_PYRKO	P58502 pyrococcus
28	36	52.9	464	1	LEU2_BACCR	Q81g10 bacillus ce
29	36	52.9	471	1	CD36_BOVIN	P26201 bos taurus
30	36	52.9	590	1	MUTL_THETN	Q8ra70 thermoanaer
31	36	52.9	609	1	GLMS_CLOPE	Q8xhz7 c glucosami
32	36	52.9	614	1	CPR1_DROME	Q9vn93 drosophila
33	36	52.9	624	1	PLB2_SCHPO	O13857 schizosacch
34	36	52.9	699	1	EFG_HAEIN	P43925 haemophilus
35	36	52.9	700	1	EFG_PASMU	P57938 pasteurella
36	36	52.9	750	1	ELS_CHICK	P07916 gallus gall
37	36	52.9	844	1	HEXA_STRPN	P10564 streptococc
38	36	52.9	926	1	PTN4_HUMAN	P29074 homo sapien
39	36	52.9	1103	1	VG37_BPAR1	Q9g0b5 bacteriopha
40	36	52.9	1173	1	ATC2_YEAST	P38929 saccharomyc
41	36	52.9	3703	1	ABF1_HUMAN	Q15911 homo sapien
42	35	51.5	100	1	KV3C_HUMAN	P01621 homo sapien
43	35	51.5	109	1	KV3B_HUMAN	P01620 homo sapien
44	35	51.5	109	1	KV3D_HUMAN	P01622 homo sapien
45	35	51.5	109	1	KV3E_HUMAN	P01623 homo sapien

ALIGNMENTS

RESULT 1
COBL_METJA
ID COBL_METJA STANDARD; PRT; 211 AA.
AC Q58917;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable precorrin-6Y C5,15-methyltransferase [decarboxylating]
DE (EC 2.1.1.132) (Precorrin-6 methyltransferase) (Precorrin-6Y
DE methylase).
GN COBL OR MJ1522.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: CATALYZES THE METHYLATION OF BOTH C-5 AND C-15 IN
CC PRECORRIN-6Y TO FORM PRECORRIN-8X (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 2 S-adenosyl-L-methionine + precorrin-6Y = 2
CC S-adenosyl-L-homocysteine + precorrin-8X + CO(2).
CC -!- PATHWAY: Cobalamin biosynthesis.
CC -!- SIMILARITY: TO S.TYPHIMURIUM CBIE; ALSO, LOW, TO OTHER
CC METHYLASES INVOLVED IN COBALAMIN BIOSYNTHESIS.
CC ---

CC ---
DR EMBL; U67593; AAB99541.1; -.
DR PIR; A64490; A64490.
DR TIGR; MJ1522; -.
DR InterPro; IPR000878; Cor/por_Metransf.
DR Pfam; PF00590; TP_methylase; 1.
KW Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW Methyltransferase; Complete proteome.
SQ SEQUENCE 211 AA; 23805 MW; 279A1A2B14369510 CRC64;

Query Match 63.2%; Score 43; DB 1; Length 211;

```
  Best Local Similarity   54.5%;  Pred. No. 3.8;
  Matches    6;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 VVGLSPGEQEY 13
              :||: ||::||
Db          4 IVGIGPGDREY 14


RESULT 2
PLB5_SCHPO
ID   PLB5_SCHPO     STANDARD;      PRT;   633 AA.
AC   Q9Y7N6;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Putative lysophospholipase C1450.09c precursor (EC 3.1.1.5)
DE   (Phospholipase B).
GN   SPCC1450.09C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: Catalyzes the release of fatty acids from
CC       lysophospholipids (By similarity).
CC   -!- CATALYTIC ACTIVITY: 2-lysophosphatidylcholine + H(2)O =
CC       glycerophosphocholine + a fatty acid anion.
CC   -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC   -!- SIMILARITY: Belongs to the lysophospholipase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL049559; CAB40176.2; -.
DR   GeneDB_SPombe; SPCC1450.09c; -.
DR   InterPro; IPR002642; PLAc.
DR   Pfam; PF01735; PLA2_B; 1.
DR   SMART; SM00022; PLAc; 1.
KW   Hypothetical protein; Lipid degradation; Hydrolase; Glycoprotein;
KW   Signal.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    633       PUTATIVE LYSOPHOSPHOLIPASE C1450.09C.
FT   CARBOHYD    118    118       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    153    153       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    232    232       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    256    256       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    264    264       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    293    293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    331    331       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    360    360       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    367    367       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    400    400       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    403    403       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    474    474       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    508    508       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    513    513       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    537    537       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    564    564       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    586    586       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    603    603       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   633 AA;  68292 MW;  49871B2955893D19 CRC64;

  Query Match             63.2%;  Score 43;  DB 1;  Length 633;
  Best Local Similarity   69.2%;  Pred. No. 11;
  Matches    9;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQEY 13
              ||  ||| ||||:
Db         76 PASDGLSTGEQEF 88


RESULT 3
RUVX_CAUCR
ID   RUVX_CAUCR     STANDARD;      PRT;   156 AA.
AC   Q9A5K8;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Putative Holliday junction resolvase (EC 3.1.-.-).
GN   CC2439.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC   Caulobacteraceae; Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC   -!- FUNCTION: Could be a nuclease that resolves Holliday junction
CC       intermediates in genetic recombination.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: Belongs to the yqgF HJR family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; AE005913; AAK24410.1; -.
DR   PIR; F87551; F87551.
DR   TIGR; CC2439; -.
DR   HAMAP; MF_00651; -; 1.
DR   InterPro; IPR005227; Cons_hypoth250.
DR   InterPro; IPR006641; YqgFc.
DR   Pfam; PF03652; UPF0081; 1.
DR   SMART; SM00732; YqgFc; 1.
DR   TIGRFAMs; TIGR00250; TIGR00250; 1.
KW   Hydrolase; Nuclease; DNA repair; DNA recombination; Complete proteome.
SQ   SEQUENCE   156 AA;  17142 MW;  21F54D8648396141 CRC64;

  Query Match             60.3%;  Score 41;  DB 1;  Length 156;
  Best Local Similarity   80.0%;  Pred. No. 6.3;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 AVVGLSPGEQ 11
              ||||| |||:
Db         18 AVVGLDPGEK 27


RESULT 4
MER_METTI
ID   MER_METTI      STANDARD;      PRT;   326 AA.
AC   Q9UXP0;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Coenzyme F420-dependent N(5),N(10)-methylenetetrahydromethanopterin
DE   reductase (EC 1.5.99.11) (Methylene-H(4)MPT reductase).
GN   MER OR FFDA.
OS   Methanolobus tindarius.
OC   Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC   Methanosarcinaceae; Methanolobus.
OX   NCBI_TaxID=2221;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 2278;
RX   MEDLINE=99132696; PubMed=9933933;
RA   Westenberg D.J., Braune A., Ruppert C., Mueller V., Herzberg C.,
RA   Gottschalk G., Blaut M.;
RT   "The F420H2-dehydrogenase from Methanolobus tindarius: cloning of the
RT   ffd operon and expression of the genes in Escherichia coli.";
RL   FEMS Microbiol. Lett. 170:389-398(1999).
CC   -!- FUNCTION: Catalyzes the reversible reduction of methylene-H(4)MPT
CC       to methyl-H(4)MPT (By similarity).
CC   -!- CATALYTIC ACTIVITY: N(5),N(10)-methylenetetrahydromethanopterin +
CC       reduced coenzyme F420 = 5-methyl-5,6,7,8-tetrahydromethanopterin +
CC       coenzyme F420.
CC   -!- PATHWAY: Methanogenesis from carbon dioxide; fifth step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the mer family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ011519; CAB56639.1; -.
DR   PIR; T45226; T45226.
DR   HAMAP; MF_01091; -; 1.
DR   InterPro; IPR002103; Bac_luciferase.
DR   Pfam; PF00296; bac_luciferase; 1.
KW   Methanogenesis; One-carbon metabolism; Oxidoreductase.
SQ   SEQUENCE   326 AA;  34043 MW;  16F3AB9733A45D82 CRC64;

  Query Match             60.3%;  Score 41;  DB 1;  Length 326;
  Best Local Similarity   70.0%;  Pred. No. 13;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 AVVGLSPGEQ 11
              |::|| ||||
Db         88 AILGLGPGEQ 97


RESULT 5
COBI_MYCTU
ID   COBI_MYCTU     STANDARD;      PRT;   508 AA.
AC   Q10677;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Cobalamin biosynthesis protein COBIJ [Includes: Precorrin-2 C20-
DE   methyltransferase (EC 2.1.1.130) (S-adenosyl-L-methionine--precorrin-2
DE   methyltransferase) (SP2MT); Precorrin-3 methylase (EC 2.1.1.-)].
GN   COBIJ OR COBI OR RV2066 OR MT2126 OR MTCY49.05 OR MB2092.
OS   Mycobacterium tuberculosis, and
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773, 1765;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX   MEDLINE=22206494; PubMed=12218036;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT   "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   J. Bacteriol. 184:5479-5490(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.bovis; STRAIN=AF2122/97;
RX   MEDLINE=22709107; PubMed=12788972;
RA   Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA   Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA   Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA   Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT   "The complete genome sequence of Mycobacterium bovis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC   -!- FUNCTION: METHYLATES PRECORRIN-2 AT THE C-20 POSITION TO PRODUCE
CC       PRECORRIN-3A (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-2 = S-
CC       adenosyl-L-homocysteine + precorrin-3A.
CC   -!- PATHWAY: Cobalamin biosynthesis.
CC   -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS SUMT, CYSG, CBIF/COBM
CC       AND CBIL/COBI.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ---------------------------------------------------------------------
DR   EMBL; Z73966; CAA98214.1; -.
DR   EMBL; AE007063; AAK46406.1; -.
DR   EMBL; BX248341; CAD96945.1; -.
DR   PIR; E70764; E70764.
DR   TIGR; MT2126; -.
DR   TubercuList; Rv2066; -.
DR   InterPro; IPR006364; CobI_CbiL.
DR   InterPro; IPR006363; CobJ.
DR   InterPro; IPR000878; Cor/por_Metransf.
DR   InterPro; IPR003043; Uropor_Metransf.
DR   Pfam; PF00590; TP_methylase; 2.
DR   TIGRFAMs; TIGR01467; cobI_cbiL; 1.
DR   TIGRFAMs; TIGR01466; cobJ_cbiH; 1.
DR   PROSITE; PS00839; SUMT_1; 1.
DR   PROSITE; PS00840; SUMT_2; 1.
KW   Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW   Methyltransferase; Multifunctional enzyme; Complete proteome.
FT   DOMAIN        1    243       PRECORRIN-2 C20-METHYLTRANSFERASE.
FT   DOMAIN      244    508       PRECORRIN-3 METHYLASE.
SQ   SEQUENCE   508 AA;  53910 MW;  95AC066F022C4DC1 CRC64;

  Query Match             60.3%;  Score 41;  DB 1;  Length 508;
  Best Local Similarity   58.3%;  Pred. No. 20;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 AVVGLSPGEQEY 13
              ||||| ||: ::
Db        250 AVVGLGPGDSDW 261


RESULT 6
PLB4_SCHPO
ID   PLB4_SCHPO     STANDARD;      PRT;   673 AA.
AC   Q9P327;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Putative lysophospholipase C977.09c precursor (EC 3.1.1.5)
DE   (Phospholipase B).
GN   SPAC977.09C OR SPAC1348.10C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: Catalyzes the release of fatty acids from
CC       lysophospholipids (By similarity).
CC   -!- CATALYTIC ACTIVITY: 2-lysophosphatidylcholine + H(2)O =
CC       glycerophosphocholine + a fatty acid anion.
CC   -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC   -!- SIMILARITY: Belongs to the lysophospholipase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL358912; CAB94277.1; -.
DR   EMBL; AL137130; CAB69631.1; -.
DR   PIR; T50281; T50281.
DR   GeneDB_SPombe; SPAC977.09c; -.
DR   InterPro; IPR001179; FKBP_PPIase.
DR   InterPro; IPR002642; PLAc.
DR   Pfam; PF01735; PLA2_B; 1.
DR   SMART; SM00022; PLAc; 1.
KW   Hypothetical protein; Lipid degradation; Hydrolase; Glycoprotein;
KW   Signal.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    673       PUTATIVE LYSOPHOSPHOLIPASE C977.09C.
FT   CARBOHYD     72     72       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    191    191       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    194    194       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    301    301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    374    374       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    404    404       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    409    409       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    481    481       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    516    516       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    545    545       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    574    574       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   673 AA;  74595 MW;  B39A773E76CD694B CRC64;

  Query Match             58.8%;  Score 40;  DB 1;  Length 673;
  Best Local Similarity   61.5%;  Pred. No. 39;
  Matches    8;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQEY 13
              ||  ||: ||| |
Db         83 PASEGLNEGEQSY 95


RESULT 7
KV3I_HUMAN
ID   KV3I_HUMAN     STANDARD;      PRT;   115 AA.
AC   P04433;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Ig kappa chain V-III region VG precursor (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85087932; PubMed=6440122;
RA   Pech M., Zachau H.G.;
RT   "Immunoglobulin genes of different subgroups are interdigitated
RT   within the VK locus.";
RL   Nucleic Acids Res. 12:9229-9236(1984).
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X01668; -; NOT_ANNOTATED_CDS.
DR   PIR; A01900; K3HUVG.
DR   HSSP; P80362; 1WTL.
DR   GO; GO:0005576; C:extracellular; NAS.
DR   GO; GO:0003823; F:antigen binding; NAS.
DR   GO; GO:0006955; P:immune response; NAS.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Immunoglobulin V region; Signal.
FT   SIGNAL        1     20
FT   CHAIN        21   >115       IG KAPPA CHAIN V-III REGION VG.
FT   DOMAIN       21     43       FRAMEWORK-1.
FT   DOMAIN       44     54       COMPLEMENTARITY-DETERMINING-1.
FT   DOMAIN       55     69       FRAMEWORK-2.
FT   DOMAIN       70     76       COMPLEMENTARITY-DETERMINING-2.
FT   DOMAIN       77    108       FRAMEWORK-3.
FT   DOMAIN      109    115       COMPLEMENTARITY-DETERMINING-3.
FT   DISULFID     43    108       BY SIMILARITY.
FT   NON_TER     115    115
SQ   SEQUENCE   115 AA;  12575 MW;  2DE47CDA3A17D555 CRC64;

  Query Match             57.4%;  Score 39;  DB 1;  Length 115;
  Best Local Similarity   63.6%;  Pred. No. 10;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQ 11
              || : |||||:
Db         28 PATLSLSPGER 38


RESULT 8
KV5I_MOUSE
ID   KV5I_MOUSE     STANDARD;      PRT;   115 AA.
AC   P01642;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Ig kappa chain V-V region L7 precursor (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=81220975; PubMed=6264318;
RA   Pech M., Hochtl J., Schnell H., Zachau H.G.;
RT   "Differences between germ-line and rearranged immunoglobulin V kappa
RT   coding sequences suggest a localized mutation mechanism.";
RL   Nature 291:668-670(1981).
CC   -!- MISCELLANEOUS: THERE APPEAR TO BE TWO POSSIBLE SPLICE JUNCTIONS AT
CC       THE 3' END OF THE INTRON. THE ALTERNATE WOULD CODE FOR A PROTEIN
CC       LACKING RESIDUES 17-19.
DR   PIR; A01925; KVMSL7.
DR   PDB; 1J10; 18-FEB-03.
DR   PDB; 1J1P; 18-FEB-03.
DR   PDB; 1J1X; 18-FEB-03.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Immunoglobulin V region; Signal; 3D-structure.
FT   SIGNAL        1     20
FT   CHAIN        21   >115       IG KAPPA CHAIN V-V REGION L7.
FT   DOMAIN       21     43       FRAMEWORK-1.
FT   DOMAIN       44     54       COMPLEMENTARITY-DETERMINING-1.
FT   DOMAIN       55     69       FRAMEWORK-2.
FT   DOMAIN       70     76       COMPLEMENTARITY-DETERMINING-2.
FT   DOMAIN       77    108       FRAMEWORK-3.
FT   DOMAIN      109   >115       COMPLEMENTARITY-DETERMINING-3.
FT   DISULFID     43    108       BY SIMILARITY.
FT   NON_TER     115    115
SQ   SEQUENCE   115 AA;  12615 MW;  C17BEC758C577E00 CRC64;

  Query Match             57.4%;  Score 39;  DB 1;  Length 115;
  Best Local Similarity   54.5%;  Pred. No. 10;
  Matches    6;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQ 11
              ||:: :||||:
Db         28 PAILSVSPGER 38


RESULT 9
BUD3_YEAST
ID   BUD3_YEAST     STANDARD;      PRT;  1636 AA.
AC   P25558; P25556; P25557; P87007;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Bud site selection protein BUD3.
GN   BUD3 OR YCL014W/YCL013W/YCL012W OR YCL14W/YCL13W/YCL12W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95247824; PubMed=7730410;
RA   Chant J., Mischke M., Mitchell E., Herskowitz I., Pringle J.R.;
RT   "Role of Bud3p in producing the axial budding pattern of yeast.";
RL   J. Cell Biol. 129:767-778(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=92244356; PubMed=1574125;
RA   Oliver S.G., van der Aart Q.J.M., Agostoni-Carbone M.L., Aigle M.,
RA   Alberghina L., Alexandraki D., Antoine G., Anwar R., Ballesta J.P.G.,
RA   Benit P., Berben G., Bergantino E., Biteau N., Bolle P.-A.,
RA   Bolotin-Fukuhara M., Brown A.J.P., Brown R., Buhler J.-M.,
RA   Carignani G., Chanet R., Contreras R., Crouzet M., Daignan-Fornier B.,
RA   De Haan M., Defoor E., Delgado M.D., Demolder J., Doira C., Dubois E.,
RA   Dujon B., Duesterhoeft A., Erdmann D., Esteban M., Fabre F.,
RA   Fairhead C.A., Faye G., Feldmann H., Fiers W.,
RA   Francingues-Gaillard M.-C., Franco L., Frontali L., Fukuhara H.,
RA   Fuller L.J., Gent M.E., Gigot D., Gilliquet V., Glansdorff N.,
RA   Goffeau A., Grenson M., Grisanti P., Grivell L.A., Haasemann M.,
RA   Hatat D., Hegemann J.H., Herbert C.J., Hilger F., Hohmann S.,
RA   Hollenberg C.P., Huse K., Iborra F., Indge K.J., Isono K., Jackman P.,
RA   Jacq C., Jacquet M., James C.M., Jauniaux J.-C., Jia Y., Jimenez A.,
RA   Kleinhans U., Kreisl P., Lafranchi G., Lewis C., van der Linden C.G.,
RA   Lucchini G., Lutzenkirchen K., Maat C., Mannhaupt G., Manzano M.E.,
RA   Martegani E., Mathieu A., Maurer C.T.C., McConnell D., McKee R.A.,
RA   Messenguy F., Mewes H.-W., Molemans F., Montague M.A., Navas L.,
RA   Newlon C.S., Olson M.V., Pallier C., Panzeri L., Pearson B.M.,
RA   Perea J., Philippsen P., Pierard A., Planta R.J., Plevani P.,
RA   Poetsch B., Pohl F.M., Purnelle B., Ramezani Rad M., Rasmussen S.W.,
RA   Raynal A., Remacha M., Richterich P., Roberts A.B., Rodriguez F.,
RA   Sanz E., Schaaff-Gerstenschlaeger I., Scherens B., Schweitzer B.,
RA   Shu Y., Skala J., Slonimski P.P., Sor F., Soustelle C.,
RA   Spiegelberg R., Stateva L.I., Steensma H.Y., Steiner S., Thierry A.,
RA   Thireos G., Triano L.N., Urrestarazu L.A., Valle G., Vetter I.,
RA   van Vliet-Reedijk J.C., Volckaert G., Vreken P., Warmington J.R.,
RA   von Wettstein D., Wicksteed B.L., Wilson C., Wurst H., Xu G.,
```

```
RA   Zimmermann F.K., Sgouros J.G.;
RT   "The complete DNA sequence of yeast chromosome III.";
RL   Nature 357:38-46(1992).
RN   [3]
RP   REVISIONS.
RA   Gromadka R.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   REVISIONS.
RA   Valles G., Volckaerts G.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Co-assembles with BUD4 at bud sites. BUD4 and BUD3 may
CC       cooperate to recognize a spatial landmark (the neck filaments)
CC       during mitosis and they subsequently become a landmark for
CC       establishing the axial budding pattern in G1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17580; AAA86315.1; -.
DR   EMBL; X59720; CAA42346.2; -.
DR   PIR; S74285; S74285.
DR   PIR; S74286; S74286.
DR   GermOnline; 138855; -.
DR   SGD; S0000520; BUD3.
DR   InterPro; IPR000219; RhoGEF.
DR   SMART; SM00325; RhoGEF; 1.
KW   Cell cycle.
SQ   SEQUENCE   1636 AA;  184717 MW;  9E4E46BA5C3A3F69 CRC64;

  Query Match             57.4%;  Score 39;  DB 1;  Length 1636;
  Best Local Similarity   58.3%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQE 12
              ||:  ||| :||
Db        112 PAIENLSPSDQE 123


RESULT 10
RS1_LEULA
ID   RS1_LEULA      STANDARD;      PRT;   429 AA.
AC   P50889; P71450;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   40S ribosomal protein S1.
GN   RPS1.
OS   Leuconostoc lactis.
OC   Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX   NCBI_TaxID=1246;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97186703; PubMed=9034319;
RA   Yamit-Hezi A., Levy Z., Neuman S., Nudel U.;
RT   "A Leuconostoc lactis protein with homology to ribosomal protein S1
RT   shares common epitopes and common DNA binding properties with a
RT   mammalian DNA binding nuclear factor.";
RL   Gene 185:99-103(1997).
RN   [2]
RP   SEQUENCE OF 24-429 FROM N.A.
RX   MEDLINE=95237615; PubMed=7721096;
RA   Eklund E.A., Lee S.W., Skalnik D.G.;
RT   "Cloning of a cDNA encoding a human DNA-binding protein similar to
RT   ribosomal protein S1.";
RL   Gene 155:231-235(1995).
RN   [3]
RP   SEQUENCE OF 78-429 FROM N.A.
RX   MEDLINE=96164600; PubMed=8568274;
RA   Tsuzaka K., Leu A.K., Frank M.B., Movafagh B.F., Koscec M.,
RA   Winkler T.H., Kalden J.R., Reichlin M.;
RT   "Lupus autoantibodies to double-stranded DNA cross-react with
RT   ribosomal protein S1.";
RL   J. Immunol. 156:1668-1675(1996).
CC   -!- FUNCTION: EXHIBITS PREFERENTIAL BINDING TO SINGLE-STRANDED AND
CC       DOUBLE-STRANDED DNA AND A LOW BINDING AFFINITY FOR RNA.
CC   -!- SIMILARITY: Belongs to the S1P family of ribosomal proteins.
CC   -!- SIMILARITY: Contains 4 S1 motif domains.
CC   -!- CAUTION: WAS ORIGINALLY (REF.2 AND REF.3) THOUGHT TO ORIGINATE
CC       FROM HUMAN BUT IS MOST PROBABLY THE RESULT OF A CDNA LIBRARY
CC       CONTAMINATION BY L.LACTIS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U24086; AAB08978.1; -.
DR   EMBL; U05589; AAA77669.1; -.
DR   EMBL; U27517; AAA97575.1; -.
DR   HSSP; P05055; 1SRO.
DR   InterPro; IPR008994; Nucleic_acid_OB.
DR   InterPro; IPR000110; Ribosomal_S1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF00575; S1; 4.
DR   PRINTS; PR00681; RIBOSOMALS1.
DR   SMART; SM00316; S1; 4.
DR   PROSITE; PS50126; S1; 4.
KW   Ribosomal protein; Repeat; RNA-binding.
FT   DOMAIN       55    128       S1 MOTIF 1.
FT   DOMAIN      144    211       S1 MOTIF 2.
FT   DOMAIN      231    299       S1 MOTIF 3.
FT   DOMAIN      316    385       S1 MOTIF 4.
FT   CONFLICT     24     24       S -> G (IN REF. 2).
FT   CONFLICT    122    122       A -> S (IN REF. 3).
FT   CONFLICT    217    217       L -> R (IN REF. 2 AND 3).
SQ   SEQUENCE   429 AA;  46386 MW;  92AC82605F39DDFC CRC64;

  Query Match             55.9%;  Score 38;  DB 1;  Length 429;
  Best Local Similarity   80.0%;  Pred. No. 55;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 AVVGLSPGEQ 11
              ||||||  ||:
Db         71 AVVGLSTGEE 80


RESULT 11
CATA_PICAN
ID   CATA_PICAN     STANDARD;      PRT;   507 AA.
AC   P30263;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Peroxisomal catalase (EC 1.11.1.6).
GN   PXP9 OR PXP-9.
OS   Pichia angusta (Yeast) (Hansenula polymorpha).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Pichia.
OX   NCBI_TaxID=4905;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 34438;
RX   MEDLINE=92299073; PubMed=1607006;
RA   Didion T., Roggenkamp R.O.;
RT   "Targeting signal of the peroxisomal catalase in the methylotrophic
RT   yeast Hansenula polymorpha.";
RL   FEBS Lett. 303:113-116(1992).
```

```
CC   -!- FUNCTION: Occurs in almost all aerobically respiring organisms and
CC       serves to protect cells from the toxic effects of hydrogen
CC       peroxide.
CC   -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC   -!- COFACTOR: Heme group.
CC   -!- SUBUNIT: Homotetramer.
CC   -!- SUBCELLULAR LOCATION: Peroxisomal.
CC   -!- SIMILARITY: Belongs to the catalase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56501; CAA39856.1; -.
DR   PIR; S23422; S23422.
DR   HSSP; P15202; 1A4E.
DR   InterPro; IPR002226; Catalase.
DR   Pfam; PF00199; catalase; 1.
DR   PRINTS; PR00067; CATALASE.
DR   ProDom; PD000510; Catalase; 1.
DR   PROSITE; PS00437; CATALASE_1; 1.
DR   PROSITE; PS00438; CATALASE_2; 1.
KW   Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
KW   Peroxisome.
FT   ACT_SITE     65     65       BY SIMILARITY.
FT   ACT_SITE    138    138       BY SIMILARITY.
FT   METAL       348    348       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT   SITE        505    507       MICROBODY TARGETING SIGNAL (POTENTIAL).
SQ   SEQUENCE   507 AA;  57849 MW;  3536ED0A49539CC3 CRC64;

  Query Match             55.9%;  Score 38;  DB 1;  Length 507;
  Best Local Similarity   70.0%;  Pred. No. 65;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 VVGLSPGEQE 12
              |:| :|||||
Db        445 VLGRTPGEQE 454


RESULT 12
COBJ_ARCFU
ID   COBJ_ARCFU     STANDARD;      PRT;   446 AA.
AC   O29534;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cobalamin biosynthesis protein cobJH [Includes: Precorrin-3B C17-
DE   methyltransferase (EC 2.1.1.131) (Precorrin-3 methyltransferase)
DE   (Precorrin-3 methylase); Precorrin-8X methylmutase (EC 5.4.1.2)
DE   (Precorrin isomerase)].
GN   COBJH OR AF0724.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC   Archaeoglobaceae; Archaeoglobus.
OX   NCBI_TaxID=2234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE=98049343; PubMed=9389475;
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA   Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA   Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA   Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA   Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA   Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA   Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA   Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA   Venter J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
CC   -!- FUNCTION: BIFUNCTIONAL ENZYME WITH A METHYLTRANSFERASE DOMAIN THAT
CC       CATALYZES THE METHYLATION OF C-17 IN PRECORRIN-3B TO FORM
CC       PRECORRIN-4 AND AN ISOMERASE DOMAIN THAT CATALYZES THE CONVERSION
CC       OF PRECORRIN-8X TO HYDROGENOBYRINIC ACID; A METHYL MIGRATION
CC       REACTION DURING THE TRANSFORMATION OF PRECORRIN-3 TO FORM
CC       COBYRINIC ACID (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-3B = S-
CC       adenosyl-L-homocysteine + precorrin 4.
CC   -!- CATALYTIC ACTIVITY: Precorrin-8X = hydrogenobyrinate.
CC   -!- PATHWAY: Cobalamin biosynthesis.
CC   -!- SIMILARITY: Belongs to the cobH family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001055; AAB90518.1; -.
DR   PIR; D69340; D69340.
DR   TIGR; AF0724; -.
DR   InterPro; IPR003722; CbiC.
DR   InterPro; IPR006363; CobJ.
DR   InterPro; IPR000878; Cor/por_Metransf.
DR   Pfam; PF02570; CbiC; 1.
DR   Pfam; PF00590; TP_methylase; 1.
DR   TIGRFAMs; TIGR01466; cobJ_cbiH; 1.
KW   Cobalamin biosynthesis; Transferase; Methyltransferase; Isomerase;
KW   Multifunctional enzyme; Complete proteome.
FT   DOMAIN        1    246       PRECORRIN-3B C-17 METHYLTRANSFERASE.
FT   DOMAIN      247    446       PRECORRIN-8X METHYLMUTASE.
SQ   SEQUENCE   446 AA;  48678 MW;  7341BCBC0998FFDA CRC64;

  Query Match             54.4%;  Score 37;  DB 1;  Length 446;
  Best Local Similarity   60.0%;  Pred. No. 85;
  Matches    6;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 VVGLSPGEQE 12
              |||: ||::|
Db         12 VVGIGPGKEE 21


RESULT 13
ENO1_MAIZE
ID   ENO1_MAIZE     STANDARD;      PRT;   446 AA.
AC   P26301;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Enolase 1 (EC 4.2.1.11) (2-phosphoglycerate dehydratase 1) (2-phospho-
DE   D-glycerate hydro-lyase 1).
GN   ENO1 OR PGH1.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Berkeley Fast; TISSUE=Root;
RX   MEDLINE=91316216; PubMed=1859865;
RA   Lal S.K., Johnson S., Conway T., Kelley P.M.;
RT   "Characterization of a maize cDNA that complements an enolase-
RT   deficient mutant of Escherichia coli.";
RL   Plant Mol. Biol. 16:787-795(1991).
CC   -!- CATALYTIC ACTIVITY: 2-phospho-D-glycerate = phosphoenolpyruvate +
CC       H(2)O.
```

```
CC   -!- COFACTOR: Magnesium is required for catalysis and for stabilizing
CC       the dimer (By similarity).
CC   -!- PATHWAY: Glycolysis.
CC   -!- SUBUNIT: Homodimer.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the enolase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X55981; CAA39454.1; -.
DR   PIR; S16257; S16257.
DR   HSSP; P56252; 1PDZ.
DR   MaizeDB; 30060; -.
DR   InterPro; IPR000941; Enolase.
DR   Pfam; PF00113; enolase; 1.
DR   Pfam; PF03952; enolase_N; 1.
DR   PRINTS; PR00148; ENOLASE.
DR   ProDom; PD000902; Enolase; 1.
DR   TIGRFAMs; TIGR01060; eno; 1.
DR   PROSITE; PS00164; ENOLASE; 1.
KW   Lyase; Glycolysis; Magnesium; Multigene family.
FT   ACT_SITE    164    164       BY SIMILARITY.
FT   METAL       251    251       MAGNESIUM (BY SIMILARITY).
FT   METAL       302    302       MAGNESIUM (BY SIMILARITY).
FT   METAL       329    329       MAGNESIUM (BY SIMILARITY).
SQ   SEQUENCE   446 AA;  48063 MW;  6266C48914F35198 CRC64;

  Query Match             54.4%;  Score 37;  DB 1;  Length 446;
  Best Local Similarity   63.6%;  Pred. No. 85;
  Matches    7;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQ 11
              ||:||  | ||
Db         77 PAIVGKDPTEQ 87


RESULT 14
NR41_RAT
ID   NR41_RAT       STANDARD;      PRT;   597 AA.
AC   P22829;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Orphan nuclear receptor HMR (Nerve growth factor induced protein I-B)
DE   (NGFI-B) (NUR77).
GN   NR4A1 OR HMR OR NGFIB.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90166506; PubMed=3272167;
RA   Milbrandt J.;
RT   "Nerve growth factor induces a gene homologous to the glucocorticoid
RT   receptor gene.";
RL   Neuron 1:183-188(1988).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=93361012; PubMed=8395013;
RA   Wilson T.E., Fahrner T.J., Milbrandt J.;
RT   "The orphan receptors NGFI-B and steroidogenic factor 1 establish
RT   monomer binding as a third paradigm of nuclear receptor-DNA
RT   interaction.";
RL   Mol. Cell. Biol. 13:5794-5804(1993).
RN   [3]
RP   DNA BINDING MOTIFS.
RX   MEDLINE=92229411; PubMed=1314418;
RA   Wilson T.E., Paulsen R.E., Padgett K.A., Milbrandt J.;
RT   "Participation of non-zinc finger residues in DNA binding by two
RT   nuclear orphan receptors.";
RL   Science 256:107-110(1992).
RN   [4]
RP   PHOSPHORYLATION.
RX   MEDLINE=94043340; PubMed=8227042;
RA   Hirata Y., Kiuchi K., Chen H.-C., Milbrandt J., Guroff G.;
RT   "The phosphorylation and DNA binding of the DNA-binding domain of the
RT   orphan nuclear receptor NGFI-B.";
RL   J. Biol. Chem. 268:24808-24812(1993).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 264-351 IN COMPLEX WITH NBRE,
RP   AND METAL-BINDING.
RX   MEDLINE=99260743; PubMed=10331876;
RA   Meinke G., Sigler P.B.;
RT   "DNA-binding mechanism of the monomeric orphan nuclear receptor
RT   NGFI-B.";
RL   Nat. Struct. Biol. 6:471-477(1999).
CC   -!- FUNCTION: Probable nuclear receptor. May act concomitantly with
CC       NURR1 in regulating the expression of delayed-early genes during
CC       liver regeneration. Binds the NGFI-B response element (NBRE) 5'-
CC       AAAAGGTCA-3'.
CC   -!- SUBUNIT: Binds DNA as a monomer.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Expressed in lung, brain and superior
CC       cervical ganglia. High levels are seen in the adrenal tissue.
CC   -!- INDUCTION: By nerve growth factor and during liver regeneration.
CC   -!- PTM: Phosphorylation of Ser-350 results in decrease in NBRE
CC       binding while phosphorylation of Ser-340 has little effect on it.
CC   -!- SIMILARITY: Belongs to the nuclear hormone receptor family. NR4
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17254; AAA56770.1; ALT_INIT.
DR   PDB; 1CIT; 26-JUN-00.
DR   TRANSFAC; T00619; -.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR008946; Str_ncl_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Phosphorylation; 3D-structure.
FT   DNA_BIND     266    331       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING      266    286       C4-TYPE.
FT   ZN_FING      302    326       C4-TYPE.
FT   DOMAIN       408    458       LIGAND-BINDING (POTENTIAL).
FT   DOMAIN        80     91       POLY-SER.
FT   DOMAIN       182    186       POLY-PRO.
FT   DOMAIN       582    585       POLY-PRO.
FT   MOD_RES      340    340       PHOSPHORYLATION (BY PKA).
FT   MOD_RES      350    350       PHOSPHORYLATION (BY PKA).
FT   MUTAGEN      340    340       S->A: LOSS OF PHOSPHORYLATION.
FT   MUTAGEN      350    350       S->A: LOSS OF PHOSPHORYLATION.
FT   MUTAGEN      345    345       R->K: DECREASED NBRE BINDING.
FT   MUTAGEN      348    348       L->V: ALMOST COMPLETE LOSS OF NBRE
FT                                 BINDING.
```

```
SQ   SEQUENCE   597 AA;  64281 MW;  9CFA987112337E53 CRC64;

  Query Match             54.4%;  Score 37;  DB 1;  Length 597;
  Best Local Similarity   50.0%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQE 12
              |  : ||||:|:
Db        424 PGFIELSPGDQD 435


RESULT 15
GLMS_CLOTE
ID   GLMS_CLOTE     STANDARD;      PRT;   607 AA.
AC   Q890U2;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Glucosamine--fructose-6-phosphate aminotransferase [isomerizing]
DE   (EC 2.6.1.16) (Hexosephosphate aminotransferase) (D-fructose-6-
DE   phosphate amidotransferase) (GFAT) (L-glutamine-D-fructose-6-phosphate
DE   amidotransferase) (Glucosamine-6-phosphate synthase).
GN   GLMS OR CTC02543.
OS   Clostridium tetani.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Massachusetts / E88;
RX   MEDLINE=22457253; PubMed=12552129;
RA   Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,
RA   Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA   Gottschalk G.;
RT   "The genome sequence of Clostridium tetani, the causative agent of
RT   tetanus disease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
CC   -!- FUNCTION: Catalyzes the first step in hexosamine metabolism,
CC       converting fructose-6P into glucosamine-6P using glutamine as a
CC       nitrogen source (By similarity).
CC   -!- CATALYTIC ACTIVITY: L-glutamine + D-fructose 6-phosphate = L-
CC       glutamate + D-glucosamine 6-phosphate.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: In the C-terminal section; belongs to the SIS family.
CC       GFAT subfamily.
CC   -!- SIMILARITY: Contains 1 type-2 glutamine amidotransferase domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE015944; AAO37003.1; ALT_INIT.
DR   HAMAP; MF_00164; -; 1.
DR   InterPro; IPR000583; GATase_2.
DR   InterPro; IPR005855; GlmS.
DR   InterPro; IPR001347; SIS.
DR   Pfam; PF00310; GATase_2; 1.
DR   Pfam; PF01380; SIS; 2.
DR   TIGRFAMs; TIGR01135; glmS; 1.
DR   PROSITE; PS00443; GATASE_TYPE_II; 1.
KW   Transferase; Aminotransferase; Glutamine amidotransferase;
KW   Complete proteome.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1    239       GLUTAMINE AMIDOTRANSFERASE.
FT   ACT_SITE      1      1       GATASE (BY SIMILARITY).
FT   ACT_SITE    602    602       ISOMERIZATION FRU-6P (BY SIMILARITY).
SQ   SEQUENCE   607 AA;  67694 MW;  F4B1CFA2EEA37948 CRC64;

  Query Match             54.4%;  Score 37;  DB 1;  Length 607;
  Best Local Similarity   61.5%;  Pred. No. 1.1e+02;
  Matches    8;  Conservative    2;  Mismatches    1;  Indels    2;  Gaps    1;

Qy          1 PAVVGLSPGEQEY 13
              | :|||  ||:||
Db        176 PLIVGL--GEEEY 186


Search completed: May  7, 2004, 12:34:33
Job time : 4.68 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:27:20 ; Search time 8.32 Seconds
(without alignments)
150.299 Million cell updates/sec

Title: US-09-786-214A-13
Perfect score: 68
Sequence: 1 PAVVGLSPGEQEY 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	63.2	211	2	A64490	precorrin-6Y methy
2	43	63.2	623	2	T40991	probable lysophosp
3	43	63.2	821	2	C84304	DNA helicase [impo
4	41	60.3	156	2	F87551	conserved hypothet
5	41	60.3	326	2	T45226	probable N5,N10-me
6	41	60.3	508	2	E70764	probable cobI prot
7	40	58.8	227	2	B90400	hypothetical prote
8	40	58.8	243	2	I54459	MHC H-2K1-k - mous
9	40	58.8	381	2	F75270	cytochrome P450 -
10	40	58.8	540	2	A75250	carboxylesterase,
11	40	58.8	661	2	G84511	hypothetical prote
12	40	58.8	673	2	T50281	probable lysophosp
13	39	57.4	43	2	S21065	Ig kappa chain V r
14	39	57.4	96	2	S45441	Ig kappa chain V r
15	39	57.4	103	2	S19975	Ig kappa chain V r
16	39	57.4	106	2	PS0070	Ig kappa chain V r
17	39	57.4	106	2	PC4282	Ig kappa chain (an
18	39	57.4	107	2	S57444	Ig kappa chain V-J
19	39	57.4	108	2	C30502	Ig kappa chain V r
20	39	57.4	108	2	G44151	Ig kappa chain V r
21	39	57.4	111	2	S23628	Ig kappa chain V r
22	39	57.4	114	2	S54905	Ig kappa chain V r
23	39	57.4	115	1	K3HUVG	Ig kappa chain pre
24	39	57.4	115	1	KVMSL7	Ig kappa chain pre
25	39	57.4	115	2	S11697	Ig kappa chain pre
26	39	57.4	116	2	B25521	Ig kappa chain pre
27	39	57.4	119	2	S41816	Ig kappa chain V r
28	39	57.4	125	2	S40344	Ig kappa chain V-J
29	39	57.4	128	2	PN0445	Ig kappa chain pre
30	39	57.4	128	2	S40379	Ig kappa chain V-J
31	39	57.4	128	2	A56701	Ig kappa chain V r
32	39	57.4	129	2	S29627	Ig kappa chain V r
33	39	57.4	129	2	S40363	Ig kappa chain - h
34	39	57.4	132	2	S05268	Ig kappa chain pre
35	39	57.4	144	2	PL0106	Ig kappa chain pre
36	39	57.4	144	2	B30502	Ig heavy chain V r
37	39	57.4	215	2	A23746	Ig kappa chain V-I
38	39	57.4	428	2	AG1304	uracil permease ho
39	39	57.4	428	2	AG1676	uracil permease ho
40	39	57.4	1367	2	S74285	BUD3 protein - yea
41	38	55.9	86	2	S20649	Ig heavy chain V r
42	38	55.9	110	2	S60591	Ig light chain var
43	38	55.9	111	2	PN0537	Ig kappa chain V r
44	38	55.9	152	2	G75184	probable transcrip
45	38	55.9	155	2	G71217	probable transcrip

ALIGNMENTS

RESULT 1
A64490
precorrin-6Y methylase homolog - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 18-Aug-2003
C;Accession: A64490
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.,
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, (
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschi
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: A64490
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-211 <BUL>
A;Cross-references: GB:U67593; GB:L77117; NID:g2826427; PIDN:AAB99541.1; PID:g1592152;
C;Genetics:
A;Map position: FOR1500322-1500957
C;Superfamily: precorrin-6Y methylase CbiE

Query Match 63.2%; Score 43; DB 2; Length 211;
Best Local Similarity 54.5%; Pred. No. 6.7;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

```
Qy          3 VVGLSPGEQEY 13
              :||: ||::||
Db          4 IVGIGPGDREY 14
```

RESULT 2
T40991
probable lysophospholipase precursor - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C;Accession: T40991
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, March 1999
A;Reference number: Z21962
A;Accession: T40991
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-623 <LYN>
A;Cross-references: EMBL:AL049559; PIDN:CAB40176.1; GSPDB:GN00068; SPDB:SPCC1450.09c
A;Experimental source: strain 972h-; cosmid c1450
C;Genetics:
A;Gene: SPDB:SPCC1450.09c
A;Map position: 3
C;Superfamily: yeast lysophospholipase

Query Match 63.2%; Score 43; DB 2; Length 623;

Best Local Similarity 69.2%; Pred. No. 21;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 PAVVGLSPGEQEY 13
              ||  ||| ||||:
Db         76 PASDGLSTGEQEF 88
```

RESULT 3
C84304
DNA helicase [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: C84304
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: C84304
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-821 <STO>
A;Cross-references: GB:AE004437; NID:g10580995; PIDN:AAG19799.1; GSPDB:GN00138
C;Genetics:
A;Gene: hel

Query Match 63.2%; Score 43; DB 2; Length 821;
Best Local Similarity 90.0%; Pred. No. 27;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          2 AVVGLSPGEQ 11
              ||||||| ||
Db        326 AVVGLSPAEQ 335
```

RESULT 4
F87551
conserved hypothetical protein CC2439 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: F87551
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: F87551
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-156 <STO>
A;Cross-references: GB:AE005673; NID:g13423984; PIDN:AAK24410.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC2439
C;Superfamily: Haemophilus influenzae conserved hypothetical protein HI0305

Query Match 60.3%; Score 41; DB 2; Length 156;
Best Local Similarity 80.0%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy          2 AVVGLSPGEQ 11
              ||||| |||:
Db         18 AVVGLDPGEK 27
```

RESULT 5
T45226
probable N5,N10-methylene-tetrahydromethanopterin reductase (F420-dependent) [imported]
C;Species: Methanolobus tindarius
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C;Accession: T45226
R;Westenberg, D.J.; Braune, A.; Ruppert, C.; Mueller, V.; Herzberg, C.; Gottschalk, G.;
submitted to the EMBL Data Library, September 1998
A;Description: The F420H2-dehydrogenase from Methanolobus tindarius: Cloning of the ffd
A;Reference number: Z22947
A;Accession: T45226
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-326 <WES>
A;Cross-references: EMBL:AJ011519; PIDN:CAB56639.1
A;Experimental source: DSM 2278
C;Genetics:
A;Gene: ffdA

Query Match 60.3%; Score 41; DB 2; Length 326;
Best Local Similarity 70.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Qy          2 AVVGLSPGEQ 11
              |::|| ||||
Db         88 AILGLGPGEQ 97
```

RESULT 6
E70764
probable cobI protein - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: E70764
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: E70764
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-508 <COL>
A;Cross-references: GB:Z73966; GB:AL123456; NID:g3261577; PIDN:CAA98214.1; PID:e246996;
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: cobI

Query Match 60.3%; Score 41; DB 2; Length 508;
Best Local Similarity 58.3%; Pred. No. 37;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```
Qy          2 AVVGLSPGEQEY 13
              ||||| ||: ::
Db        250 AVVGLGPGDSDW 261
```

RESULT 7
B90400
hypothetical protein cbiE [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 18-Aug-2003
C;Accession: B90400
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, I
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: B90400
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-227 <KUR>
A;Cross-references: GB:AE006641; NID:g13815599; PIDN:AAK42457.1; GSPDB:GN00155
C;Genetics:
A;Gene: cbiE

C;Superfamily: precorrin-6Y methylase CbiE

Query Match 58.8%; Score 40; DB 2; Length 227;
Best Local Similarity 54.5%; Pred. No. 24;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVGLSPGEQEY 13
:|| : || : ||
Db 10 IVGVGPGDPEY 20

RESULT 8
I54459
MHC H-2K1-k - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C;Accession: I54459
R;Watts, S.; Davis, A.C.; Goodenow, R.S.
Immunogenetics 29, 355-357, 1989
A;Title: Sequence analysis of the C3H H-2K1-k gene: Relationship to the H-2 loci.
A;Reference number: I54459; MUID:89233303; PMID:2714856
A;Accession: I54459
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-243 <RES>
A;Cross-references: GB:M27134; NID:g199435; PIDN:AAA39610.1; PID:g387456
C;Genetics:
A;Introns: 22/1; 112/1

Query Match 58.8%; Score 40; DB 2; Length 243;
Best Local Similarity 77.8%; Pred. No. 26;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 4 VGLSPGEQE 12
:|||||:|
Db 216 LGLSPGEEE 224

RESULT 9
F75270
cytochrome P450 - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: F75270
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: F75270
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-381 <WHI>
A;Cross-references: GB:AE002076; GB:AE000513; NID:g6460285; PIDN:AAF12016.1; PID:g646029
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2473
A;Map position: 1

Query Match 58.8%; Score 40; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PAVVGLSP 8
||||||||
Db 53 PAVVGLSP 60

RESULT 10
A75250
carboxylesterase, type B - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: A75250
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: A75250
A;Molecule type: DNA
A;Residues: 1-540 <WHI>
A;Cross-references: GB:AE002092; GB:AE000513; NID:g6460455; PIDN:AAF12163.1; PID:g64604
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2626
A;Map position: 1
C;Superfamily: cholinesterase; cholinesterase homology

Query Match 58.8%; Score 40; DB 2; Length 540;
Best Local Similarity 70.0%; Pred. No. 59;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PAVVGLSPGE 10
| |:||:|||
Db 512 PQVLGLAPGE 521

RESULT 11
G84511
hypothetical protein At2g13900 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: G84511
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: G84511
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-661 <STO>
A;Cross-references: GB:AE002093; NID:g6598598; PIDN:AAF18650.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g13900
A;Map position: 2

Query Match 58.8%; Score 40; DB 2; Length 661;
Best Local Similarity 53.8%; Pred. No. 72;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PAVVGLSPGEQEY 13
| : : ||||||
Db 437 PLTLKIVPGEQEY 449

RESULT 12
T50281
probable lysophospholipase (EC 3.1.1.5) precursor SPAC977.09c [similarity] - fission yea
C;Species: Schizosaccharomyces pombe
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 19-Jan-2001
C;Accession: T50281; T42738
R;Zimmermann, W.; Wambutt, R.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, January 2000
A;Reference number: Z25053
A;Accession: T50281
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-673 <ZIM>
A;Cross-references: EMBL:AL137130; NID:g6742151; PIDN:CAB69631.1; PID:g6742159; GSPDB:GN
A;Experimental source: strain 972h(-); cosmid c977
R;Yoshioka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.

DNA Res. 4, 363-369, 1997
A;Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
A;Reference number: Z17323; MUID:98162722; PMID:9501991
A;Accession: T42738
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 427-673 <YOS>
A;Cross-references: EMBL:D89183; NID:g1749573; PIDN:BAA13845.1; PID:g1749574
A;Experimental source: strain PR745
C;Genetics:
A;Gene: SPDB:SPAC977.09c
A;Map position: 1
A;Introns: 651/3
C;Function:
A;Description: catalyzes the hydolysis of 2-lysophosphatidylcholine to glycerophosphocho
C;Superfamily: yeast lysophospholipase
C;Keywords: carboxylic ester hydrolase

Query Match 58.8%; Score 40; DB 2; Length 673;
Best Local Similarity 61.5%; Pred. No. 74;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PAVVGLSPGEQEY 13
|| || :|| |
Db 83 PASEGLNEGEQSY 95

RESULT 13
S21065
Ig kappa chain V region (anti-RH(D)) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 22-Nov-1993 #sequence_revision 24-May-1996 #text_change 09-May-1997
C;Accession: S21065
R;Dlouha, A.; Lecroisey, A.; Henschen, A.; Rouger, P.; Keil, B.
Protein Seq. Data Anal. 4, 317-318, 1991
A;Title: Subgroup assignement of a human monoclonal anti-Rh(D) antibody.
A;Reference number: S21065; MUID:92253544; PMID:1812483
A;Accession: S21065
A;Molecule type: protein
A;Residues: 1-43 <DLO>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin

Query Match 57.4%; Score 39; DB 2; Length 43;
Best Local Similarity 63.6%; Pred. No. 6.3;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PAVVGLSPGEQ 11
|| : |||||:
Db 8 PATLSLSPGER 18

RESULT 14
S45441
Ig kappa chain V region - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 21-Jan-2000
C;Accession: S45441
R;Cox, J.P.L.; Tomlinson, I.M.; Winter, G.
Eur. J. Immunol. 24, 827-836, 1994
A;Title: A directory of human germ-line V(kappa) segments reveals a strong bias in their
A;Reference number: S45324; MUID:94200218; PMID:8149953
A;Accession: S45441
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-96 <COX>
A;Cross-references: EMBL:Z27500
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin
F;16-91/Domain: immunoglobulin homology <IMM>

Query Match 57.4%; Score 39; DB 2; Length 96;
Best Local Similarity 63.6%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PAVVGLSPGEQ 11
|| : |||||:
Db 8 PATLSLSPGER 18

RESULT 15
S19975
Ig kappa chain V region (M-T408) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 21-Jan-2000
C;Accession: S19975
R;Weissenhorn, W.; Riethmueller, G.; Weiss, E.M.; Rieber, E.P.
submitted to the EMBL Data Library, March 1992
A;Description: Structural characterization of CD4 mAb.
A;Reference number: S19963
A;Accession: S19975
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-103 <WEI>
A;Cross-references: EMBL:X65097; NID:g52296; PIDN:CAA46225.1; PID:g52297
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin
F;11-85/Domain: immunoglobulin homology <IMM>

Query Match 57.4%; Score 39; DB 2; Length 103;
Best Local Similarity 63.6%; Pred. No. 16;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PAVVGLSPGEQ 11
|| : |||||:
Db 3 PATLSLSPGER 13

Search completed: May 7, 2004, 12:39:07
Job time : 9.48667 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:27:20 ; Search time 9.6 Seconds
(without alignments)
150.299 Million cell updates/sec

Title: US-09-786-214A-15
Perfect score: 75
Sequence: 1 AGLPAVVGLSPGEQE 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	50	66.7	381	2	F75270	cytochrome P450 -
2	43	57.3	659	1	GKBPT4	gene 18 protein -
3	43	57.3	659	2	A60885	tail sheath protei
4	43	57.3	821	2	C84304	DNA helicase [impo
5	42	56.0	326	2	T45226	probable N5,N10-me
6	42	56.0	508	2	E70764	probable cobI prot
7	41	54.7	156	2	F87551	conserved hypothet
8	41	54.7	164	2	A87399	ISCc1, transposase
9	41	54.7	177	2	B72580	hypothetical prote
10	41	54.7	214	2	A83416	hypothetical prote
11	40	53.3	130	2	D72783	hypothetical prote
12	40	53.3	243	2	I54459	MHC H-2K1-k - mous
13	40	53.3	358	1	JQ1870	AL1 protein - toma
14	40	53.3	359	2	S58167	fructose-bisphosph
15	40	53.3	359	2	T48396	fructose-bisphosph
16	40	53.3	428	2	AG1304	uracil permease ho
17	40	53.3	428	2	AG1676	uracil permease ho
18	40	53.3	540	2	A75250	carboxylesterase,
19	40	53.3	563	2	T36580	hypothetical prote
20	40	53.3	623	2	T40991	probable lysophosp
21	40	53.3	740	1	B65136	yhgF protein - Esc
22	40	53.3	740	2	A98160	hypothetical prote
23	40	53.3	740	2	H86005	hypothetical prote
24	40	53.3	776	2	AB0998	probable transcrip
25	40	53.3	1340	2	A39808	proteoglycan core
26	40	53.3	1506	2	JC5985	phosphoinositide 3
27	40	53.3	2327	2	T42630	aggrecan - bovine
28	39	52.0	43	2	S21065	Ig kappa chain V r
29	39	52.0	96	2	S45441	Ig kappa chain V r
30	39	52.0	103	2	S19975	Ig kappa chain V r
31	39	52.0	106	2	PS0070	Ig kappa chain V r
32	39	52.0	106	2	PC4282	Ig kappa chain (an
33	39	52.0	107	2	S57444	Ig kappa chain V-J
34	39	52.0	108	2	C30502	Ig kappa chain V r
35	39	52.0	108	2	S33988	Ig kappa chain V r
36	39	52.0	108	2	G44151	Ig kappa chain V r
37	39	52.0	111	2	S23628	Ig kappa chain V r
38	39	52.0	114	2	S54905	Ig kappa chain V r
39	39	52.0	115	1	K3HUVG	Ig kappa chain pre
40	39	52.0	115	1	KVMSL7	Ig kappa chain pre
41	39	52.0	115	2	S11697	Ig kappa chain pre
42	39	52.0	115	2	G72587	hypothetical prote
43	39	52.0	116	2	B25521	Ig kappa chain pre
44	39	52.0	119	2	S41816	Ig kappa chain V r
45	39	52.0	125	2	S40344	Ig kappa chain V-J

ALIGNMENTS

RESULT 1
F75270
cytochrome P450 - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: F75270
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; , M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: F75270
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-381 <WHI>
A;Cross-references: GB:AE002076; GB:AE000513; NID:g6460285; PIDN:AAF12016.1; PID:g64602
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2473
A;Map position: 1

Query Match 66.7%; Score 50; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          2 GLPAVVGLSP 11
              ||||||||||
Db         51 GLPAVVGLSP 60
```

RESULT 2
GKBPT4
gene 18 protein - phage T4
N;Alternate names: tail sheath protein gp18
C;Species: phage T4
A;Note: host Escherichia coli
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 23-Jul-1999
C;Accession: JF0021
R;Arisaka, F.; Nakako, T.; Takahashi, H.; Ishii, S.
J. Virol. 62, 1186-1193, 1988
A;Title: Nucleotide sequence of the tail sheath gene of bacteriophage T4 and amino acid
A;Reference number: JF0021; MUID:88155753; PMID:2964531
A;Accession: JF0021
A;Molecule type: DNA
A;Residues: 1-659 <ARI>
A;Cross-references: GB:M19085; EMBL:M36959; NID:g215949; PIDN:AAA32541.1; PID:g215950
C;Genetics:
A;Gene: 18
A;Map position: 97.352-99.328
C;Superfamily: phage T4 gene 18 protein

```
  Query Match             57.3%;  Score 43;  DB 1;  Length 659;
  Best Local Similarity   66.7%;  Pred. No. 56;
  Matches    8;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 GLPAVVGLSPGE 13
              |:| || | |||
Db        218 GIPGVVALYPGE 229
```

RESULT 3
A60885
tail sheath protein - phage T4
C;Species: phage T4
C;Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 03-May-1996
C;Accession: A60885
R;Arisaka, F.; Nakako, T.; Kumazaki, T.; Ishii, S.
J. Protein Chem. 6, 245-251, 1987
A;Title: Primary structure of the tail sheath protein of bacteriophage T4 and its gene.
A;Reference number: A60885
A;Accession: A60885
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-659 <ARI>
C;Superfamily: phage T4 gene 18 protein

```
  Query Match             57.3%;  Score 43;  DB 2;  Length 659;
  Best Local Similarity   66.7%;  Pred. No. 56;
  Matches    8;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 GLPAVVGLSPGE 13
              |:| || | |||
Db        218 GIPGVVALYPGE 229
```

RESULT 4
C84304
DNA helicase [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: C84304
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
 Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: C84304
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-821 <STO>
A;Cross-references: GB:AE004437; NID:g10580995; PIDN:AAG19799.1; GSPDB:GN00138
C;Genetics:
A;Gene: hel

```
  Query Match             57.3%;  Score 43;  DB 2;  Length 821;
  Best Local Similarity   90.0%;  Pred. No. 70;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          5 AVVGLSPGEQ 14
              |||||||.||
Db        326 AVVGLSPAEQ 335
```

RESULT 5
T45226
probable N5,N10-methylene-tetrahydromethanopterin reductase (F420-dependent) [imported]
C;Species: Methanolobus tindarius
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C;Accession: T45226
R;Westenberg, D.J.; Braune, A.; Ruppert, C.; Mueller, V.; Herzberg, C.; Gottschalk, G.;
submitted to the EMBL Data Library, September 1998
A;Description: The F420H2-dehydrogenase from Methanolobus tindarius: Cloning of the ffd
A;Reference number: Z22947
A;Accession: T45226
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-326 <WES>
A;Cross-references: EMBL:AJ011519; PIDN:CAB56639.1
A;Experimental source: DSM 2278
C;Genetics:
A;Gene: ffdA

```
  Query Match             56.0%;  Score 42;  DB 2;  Length 326;
  Best Local Similarity   57.1%;  Pred. No. 40;
  Matches    8;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 AGLPAVVGLSPGEQ 14
              :|  |::|| ||||
Db         84 SGGRAILGLGPGEQ 97
```

RESULT 6
E70764
probable cobI protein - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: E70764
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: E70764
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-508 <COL>
A;Cross-references: GB:Z73966; GB:AL123456; NID:g3261577; PIDN:CAA98214.1; PID:e246996;
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: cobI

```
  Query Match             56.0%;  Score 42;  DB 2;  Length 508;
  Best Local Similarity   57.1%;  Pred. No. 62;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 GLPAVVGLSPGEQE 15
              |  ||||| ||: :
Db        247 GTVAVVGLGPGDSD 260
```

RESULT 7
F87551
conserved hypothetical protein CC2439 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: F87551
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolo
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: F87551
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-156 <STO>
A;Cross-references: GB:AE005673; NID:g13423984; PIDN:AAK24410.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC2439
C;Superfamily: Haemophilus influenzae conserved hypothetical protein HI0305

```
  Query Match             54.7%;  Score 41;  DB 2;  Length 156;
  Best Local Similarity   80.0%;  Pred. No. 28;
```

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy          5 AVVGLSPGEQ 14
              ||||| |||:
Db         18 AVVGLDPGEK 27
```

RESULT 8
A87399
ISCc1, transposase OrfB [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 18-Jul-2001
C;Accession: A87399; E87402; E87502; E87570
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: A87399
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-164 <STO>
A;Cross-references: GB:AE005673; NID:g13422533; PIDN:AAK23189.1; GSPDB:GN00148
A;Accession: E87402
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-164 <ST2>
A;Cross-references: GB:AE005673; NID:g13422563; PIDN:AAK23217.1; GSPDB:GN00148
A;Accession: E87502
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-164 <ST3>
A;Cross-references: GB:AE005673; NID:g13423519; PIDN:AAK24017.1; GSPDB:GN00148
A;Accession: E87570
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-164 <ST4>
A;Cross-references: GB:AE005673; NID:g13424163; PIDN:AAK24561.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1206; CC1236; CC2043; CC2592
C;Superfamily: Streptomyces coelicolor probable transposase SC6G9.35c

Query Match 54.7%; Score 41; DB 2; Length 164;
Best Local Similarity 53.8%; Pred. No. 29;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```
Qy          2 GLPAVVGLSPGEQ 14
              |||  : | ||:|
Db         49 GLPVRIALGPGQQ 61
```

RESULT 9
B72580
hypothetical protein APE1922 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: B72580
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: B72580
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-177 <KAW>
A;Cross-references: DDBJ:AP000062; NID:g5105244; PIDN:BAA80927.1; PID:d1044713; PID:g510
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1922

Query Match 54.7%; Score 41; DB 2; Length 177;
Best Local Similarity 53.8%; Pred. No. 31;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```
Qy          2 GLPAVVGLSPGEQ 14
              |:|||:|| | ::
Db        151 GVPAVLGLEPWDR 163
```

RESULT 10
A83416
hypothetical protein PA1825 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A83416
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: A83416
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-214 <STO>
A;Cross-references: GB:AE004609; GB:AE004091; NID:g9947810; PIDN:AAG05214.1; GSPDB:GN00
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA1825

Query Match 54.7%; Score 41; DB 2; Length 214;
Best Local Similarity 63.6%; Pred. No. 38;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```
Qy          2 GLPAVVGLSPG 12
              || |:|:|:||
Db        113 GLAALIGLAPG 123
```

RESULT 11
D72783
hypothetical protein APE0253 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
C;Accession: D72783
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takal
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropy
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: D72783
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-130 <KAW>
A;Cross-references: DDBJ:AP000058; NID:g5103388; PIDN:BAA79166.1; PID:d1042942; PID:g51
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE0253
C;Superfamily: Aeropyrum pernix hypothetical protein APE0253

Query Match 53.3%; Score 40; DB 2; Length 130;
Best Local Similarity 58.3%; Pred. No. 33;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```
Qy          2 GLPAVVGLSPGE 13
              |:|| : | |||
Db          5 GVPAALALHPGE 16
```

RESULT 12
I54459
MHC H-2K1-k - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999

C;Accession: I54459
R;Watts, S.; Davis, A.C.; Goodenow, R.S.
Immunogenetics 29, 355-357, 1989
A;Title: Sequence analysis of the C3H H-2K1-k gene: Relationship to the H-2 loci.
A;Reference number: I54459; MUID:89233303; PMID:2714856
A;Accession: I54459
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-243 <RES>
A;Cross-references: GB:M27134; NID:g199435; PIDN:AAA39610.1; PID:g387456
C;Genetics:
A;Introns: 22/1; 112/1

Query Match 53.3%; Score 40; DB 2; Length 243;
Best Local Similarity 77.8%; Pred. No. 62;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 7 VGLSPGEQE 15
:|||||:|
Db 216 LGLSPGEEE 224

RESULT 13
JQ1870
AL1 protein - tomato mottle virus (isolate Florida)
C;Species: tomato mottle virus
C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 07-May-1999
C;Accession: JQ1870
R;Abouzid, A.M.; Polston, J.E.; Hiebert, E.
J. Gen. Virol. 73, 3225-3229, 1992
A;Title: The nucleotide sequence of tomato mottle virus, a new geminivirus isolated from
A;Reference number: JQ1869; MUID:93107858; PMID:1469361
A;Accession: JQ1870
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-358 <ABO>
A;Cross-references: GB:L14460
C;Genetics:
A;Map position: segment A
C;Superfamily: tomato golden mosaic virus AL1 protein

Query Match 53.3%; Score 40; DB 1; Length 358;
Best Local Similarity 58.3%; Pred. No. 92;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GLPAVVGLSPGE 13
|:||:| :|||
Db 296 GIPAIVLCNPGE 307

RESULT 14
S58167
fructose-bisphosphate aldolase (EC 4.1.2.13) - garden pea
N;Alternate names: fructose-1,6-bisphosphate aldolase
C;Species: Pisum sativum (garden pea)
C;Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
C;Accession: S58167
R;Pelzer-Reith, B.; Schnarrenberger, C.
submitted to the EMBL Data Library, July 1995
A;Description: Characterization of cDNA clones and expression of two cytosolic fructose-
A;Reference number: S58167
A;Accession: S58167
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-359 <PEL>
A;Cross-references: EMBL:X89829; NID:g927504; PIDN:CAA61947.1; PID:g927505
C;Superfamily: fructose-bisphosphate aldolase
C;Keywords: aldehyde-lyase; carbon-carbon lyase

Query Match 53.3%; Score 40; DB 2; Length 359;
Best Local Similarity 60.0%; Pred. No. 92;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AGLPAVVGLSPGEQE 15
| :|||| || |: |
Db 258 AAVPAVVFLSGGQSE 272

RESULT 15
T48396
fructose-bisphosphate aldolase-like protein - Arabidopsis thaliana
N;Alternate names: protein F17C15.110
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
C;Accession: T48396
R;Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K.
submitted to the Protein Sequence Database, March 2000
A;Reference number: Z24492
A;Accession: T48396
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-359 <BEV>
A;Cross-references: EMBL:AL162506
A;Experimental source: cultivar Columbia; BAC clone F17C15
C;Genetics:
A;Map position: 5
A;Introns: 10/1; 86/1
A;Note: F17C15.110
C;Superfamily: fructose-bisphosphate aldolase

Query Match 53.3%; Score 40; DB 2; Length 359;
Best Local Similarity 60.0%; Pred. No. 92;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AGLPAVVGLSPGEQE 15
| :|||| || |: |
Db 258 AAVPAVVFLSGGQSE 272

Search completed: May 7, 2004, 12:39:07
Job time : 9.76667 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:20:30 ; Search time 4.68 Seconds
(without alignments)
144.639 Million cell updates/sec

Title: US-09-786-214A-14
Perfect score: 65
Sequence: 1 LPAVVGLSPGEQE 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	41	63.1	156	1	RUVX_CAUCR	Q9a5k8 caulobacter
2	41	63.1	326	1	MER_METTI	Q9uxp0 methanolobu
3	40	61.5	633	1	PLB5_SCHPO	Q9y7n6 schizosacch
4	39	60.0	115	1	KV3I_HUMAN	P04433 homo sapien
5	39	60.0	115	1	KV5I_MOUSE	P01642 mus musculu
6	39	60.0	508	1	COBI_MYCTU	Q10677 mycobacteri
7	39	60.0	597	1	NR41_RAT	P22829 rattus norv
8	39	60.0	1402	1	N160_MOUSE	Q9z0w3 mus musculu
9	39	60.0	1636	1	BUD3_YEAST	P25558 saccharomyc
10	38	58.5	390	1	COBL_MYCTU	Q10671 mycobacteri
11	38	58.5	429	1	RS1_LEULA	P50889 leuconostoc
12	38	58.5	507	1	CATA_PICAN	P30263 pichia angu
13	38	58.5	516	1	C4AD_DROME	Q9v4t3 drosophila
14	38	58.5	699	1	EFG_HAEIN	P43925 haemophilus
15	38	58.5	700	1	EFG_PASMU	P57938 pasteurella
16	37.5	57.7	827	1	M4K1_MOUSE	P70218 mus musculu
17	37	56.9	280	1	PMXA_MOUSE	Q62066 mus musculu
18	37	56.9	281	1	PMXA_RAT	Q62782 rattus norv
19	37	56.9	446	1	COBJ_ARCFU	O29534 a cobalamin
20	37	56.9	446	1	ENO1_MAIZE	P26301 zea mays (m
21	37	56.9	557	1	PUR6_VIGAC	P55195 vigna aconi
22	37	56.9	637	1	MUTL_CAUCR	Q9rp66 caulobacter
23	37	56.9	658	1	VG18_BPT4	P13332 bacteriopha
24	37	56.9	813	1	CADM_MOUSE	Q9wtp5 mus musculu
25	37	56.9	985	1	4ET_HUMAN	Q9nra8 homo sapien
26	37	56.9	2269	1	RRPL_SV41	P35341 simian viru
27	36.5	56.2	374	1	RGSK_BOVIN	P79348 bos taurus
28	36	55.4	129	1	KV3H_HUMAN	P04207 homo sapien
29	36	55.4	211	1	COBL_METJA	Q58917 methanococc
30	36	55.4	279	1	TYSY_CAUCR	Q9a6h0 caulobacter
31	36	55.4	342	1	HUPK_AZOVI	P31878 azotobacter
32	36	55.4	348	1	HOXV_AZOVI	P40597 azotobacter
33	36	55.4	359	1	ALF2_PEA	P46257 pisum sativ
34	36	55.4	359	1	ALF_CICAR	O65735 cicer ariet
35	36	55.4	590	1	MUTL_THETN	Q8ra70 thermoanaer
36	36	55.4	598	1	NR41_CANFA	P51666 canis famil
37	36	55.4	750	1	ELS_CHICK	P07916 gallus gall
38	35.5	54.6	833	1	M4K1_HUMAN	Q92918 homo sapien
39	35	53.8	100	1	KV3C_HUMAN	P01621 homo sapien
40	35	53.8	109	1	KV3B_HUMAN	P01620 homo sapien
41	35	53.8	109	1	KV3D_HUMAN	P01622 homo sapien
42	35	53.8	109	1	KV3E_HUMAN	P01623 homo sapien
43	35	53.8	109	1	KV3G_HUMAN	P04206 homo sapien
44	35	53.8	129	1	KV3L_HUMAN	P18135 homo sapien
45	35	53.8	129	1	KV3M_HUMAN	P18136 homo sapien

ALIGNMENTS

RESULT 1
RUVX_CAUCR
ID RUVX_CAUCR STANDARD; PRT; 156 AA.
AC Q9A5K8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative Holliday junction resolvase (EC 3.1.-.-).
GN CC2439.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -!- FUNCTION: Could be a nuclease that resolves Holliday junction
CC intermediates in genetic recombination.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: Belongs to the yqgF HJR family.
CC ----------

CC ----------
DR EMBL; AE005913; AAK24410.1; -.
DR PIR; F87551; F87551.
DR TIGR; CC2439; -.
DR HAMAP; MF_00651; -; 1.
DR InterPro; IPR005227; Cons_hypoth250.
DR InterPro; IPR006641; YqgFc.
DR Pfam; PF03652; UPF0081; 1.
DR SMART; SM00732; YqgFc; 1.
DR TIGRFAMs; TIGR00250; TIGR00250; 1.
KW Hydrolase; Nuclease; DNA repair; DNA recombination; Complete proteome.
SQ SEQUENCE 156 AA; 17142 MW; 21F54D8648396141 CRC64;

Query Match 63.1%; Score 41; DB 1; Length 156;
Best Local Similarity 80.0%; Pred. No. 6.8;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 AVVGLSPGEQ 12

```
                |||||  |||:
Db          18 AVVGLDPGEK 27


RESULT 2
MER_METTI
ID   MER_METTI      STANDARD;      PRT;   326 AA.
AC   Q9UXP0;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Coenzyme F420-dependent N(5),N(10)-methylenetetrahydromethanopterin
DE   reductase (EC 1.5.99.11) (Methylene-H(4)MPT reductase).
GN   MER OR FFDA.
OS   Methanolobus tindarius.
OC   Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC   Methanosarcinaceae; Methanolobus.
OX   NCBI_TaxID=2221;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 2278;
RX   MEDLINE=99132696; PubMed=9933933;
RA   Westenberg D.J., Braune A., Ruppert C., Mueller V., Herzberg C.,
RA   Gottschalk G., Blaut M.;
RT   "The F420H2-dehydrogenase from Methanolobus tindarius: cloning of the
RT   ffd operon and expression of the genes in Escherichia coli.";
RL   FEMS Microbiol. Lett. 170:389-398(1999).
CC   -!- FUNCTION: Catalyzes the reversible reduction of methylene-H(4)MPT
CC       to methyl-H(4)MPT (By similarity).
CC   -!- CATALYTIC ACTIVITY: N(5),N(10)-methylenetetrahydromethanopterin +
CC       reduced coenzyme F420 = 5-methyl-5,6,7,8-tetrahydromethanopterin +
CC       coenzyme F420.
CC   -!- PATHWAY: Methanogenesis from carbon dioxide; fifth step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the mer family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ011519; CAB56639.1; -.
DR   PIR; T45226; T45226.
DR   HAMAP; MF_01091; -; 1.
DR   InterPro; IPR002103; Bac_luciferase.
DR   Pfam; PF00296; bac_luciferase; 1.
KW   Methanogenesis; One-carbon metabolism; Oxidoreductase.
SQ   SEQUENCE   326 AA;  34043 MW;  16F3AB9733A45D82 CRC64;

  Query Match             63.1%;  Score 41;  DB 1;  Length 326;
  Best Local Similarity   70.0%;  Pred. No. 14;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AVVGLSPGEQ 12
              |::|| ||||
Db         88 AILGLGPGEQ 97


RESULT 3
PLB5_SCHPO
ID   PLB5_SCHPO     STANDARD;      PRT;   633 AA.
AC   Q9Y7N6;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Putative lysophospholipase C1450.09c precursor (EC 3.1.1.5)
DE   (Phospholipase B).
GN   SPCC1450.09C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: Catalyzes the release of fatty acids from
CC       lysophospholipids (By similarity).
CC   -!- CATALYTIC ACTIVITY: 2-lysophosphatidylcholine + H(2)O =
CC       glycerophosphocholine + a fatty acid anion.
CC   -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC   -!- SIMILARITY: Belongs to the lysophospholipase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL049559; CAB40176.2; -.
DR   GeneDB_SPombe; SPCC1450.09c; -.
DR   InterPro; IPR002642; PLAc.
DR   Pfam; PF01735; PLA2_B; 1.
DR   SMART; SM00022; PLAc; 1.
KW   Hypothetical protein; Lipid degradation; Hydrolase; Glycoprotein;
KW   Signal.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    633       PUTATIVE LYSOPHOSPHOLIPASE C1450.09C.
FT   CARBOHYD    118    118       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    153    153       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    232    232       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    256    256       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    264    264       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    293    293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    331    331       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    360    360       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    367    367       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    400    400       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    403    403       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    474    474       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    508    508       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    513    513       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    537    537       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD     564     564       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     586     586       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     603     603       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   633 AA;  68292 MW;  49871B2955893D19 CRC64;

  Query Match             61.5%;  Score 40;  DB 1;  Length 633;
  Best Local Similarity   75.0%;  Pred. No. 39;
  Matches    9;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 PAVVGLSPGEQE 13
              ||  ||| ||||
Db         76 PASDGLSTGEQE 87


RESULT 4
KV3I_HUMAN
ID   KV3I_HUMAN     STANDARD;      PRT;   115 AA.
AC   P04433;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Ig kappa chain V-III region VG precursor (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85087932; PubMed=6440122;
RA   Pech M., Zachau H.G.;
RT   "Immunoglobulin genes of different subgroups are interdigitated
RT   within the VK locus.";
RL   Nucleic Acids Res. 12:9229-9236(1984).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X01668; -; NOT_ANNOTATED_CDS.
DR   PIR; A01900; K3HUVG.
DR   HSSP; P80362; 1WTL.
DR   GO; GO:0005576; C:extracellular; NAS.
DR   GO; GO:0003823; F:antigen binding; NAS.
DR   GO; GO:0006955; P:immune response; NAS.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Immunoglobulin V region; Signal.
FT   SIGNAL       1     20
FT   CHAIN       21   >115       IG KAPPA CHAIN V-III REGION VG.
FT   DOMAIN      21     43       FRAMEWORK-1.
FT   DOMAIN      44     54       COMPLEMENTARITY-DETERMINING-1.
FT   DOMAIN      55     69       FRAMEWORK-2.
FT   DOMAIN      70     76       COMPLEMENTARITY-DETERMINING-2.
FT   DOMAIN      77    108       FRAMEWORK-3.
FT   DOMAIN     109    115       COMPLEMENTARITY-DETERMINING-3.
FT   DISULFID    43    108       BY SIMILARITY.
FT   NON_TER    115    115
SQ   SEQUENCE   115 AA;  12575 MW;  2DE47CDA3A17D555 CRC64;

  Query Match             60.0%;  Score 39;  DB 1;  Length 115;
  Best Local Similarity   63.6%;  Pred. No. 11;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 PAVVGLSPGEQ 12
              || : |||||:
Db         28 PATLSLSPGER 38


RESULT 5
KV5I_MOUSE
ID   KV5I_MOUSE     STANDARD;      PRT;   115 AA.
AC   P01642;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Ig kappa chain V-V region L7 precursor (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=81220975; PubMed=6264318;
RA   Pech M., Hochtl J., Schnell H., Zachau H.G.;
RT   "Differences between germ-line and rearranged immunoglobulin V kappa
RT   coding sequences suggest a localized mutation mechanism.";
RL   Nature 291:668-670(1981).
CC   -!- MISCELLANEOUS: THERE APPEAR TO BE TWO POSSIBLE SPLICE JUNCTIONS AT
CC       THE 3' END OF THE INTRON. THE ALTERNATE WOULD CODE FOR A PROTEIN
CC       LACKING RESIDUES 17-19.
DR   PIR; A01925; KVMSL7.
DR   PDB; 1J1O; 18-FEB-03.
DR   PDB; 1J1P; 18-FEB-03.
DR   PDB; 1J1X; 18-FEB-03.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Immunoglobulin V region; Signal; 3D-structure.
FT   SIGNAL       1     20
FT   CHAIN       21   >115       IG KAPPA CHAIN V-V REGION L7.
FT   DOMAIN      21     43       FRAMEWORK-1.
FT   DOMAIN      44     54       COMPLEMENTARITY-DETERMINING-1.
FT   DOMAIN      55     69       FRAMEWORK-2.
FT   DOMAIN      70     76       COMPLEMENTARITY-DETERMINING-2.
FT   DOMAIN      77    108       FRAMEWORK-3.
FT   DOMAIN     109   >115       COMPLEMENTARITY-DETERMINING-3.
FT   DISULFID    43    108       BY SIMILARITY.
FT   NON_TER    115    115
SQ   SEQUENCE   115 AA;  12615 MW;  C17BEC758C577E00 CRC64;

  Query Match             60.0%;  Score 39;  DB 1;  Length 115;
  Best Local Similarity   54.5%;  Pred. No. 11;
  Matches    6;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 PAVVGLSPGEQ 12
              ||:: :||||:
Db         28 PAILSVSPGER 38


RESULT 6
COBI_MYCTU
ID   COBI_MYCTU     STANDARD;      PRT;   508 AA.
AC   Q10677;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Cobalamin biosynthesis protein COBIJ [Includes: Precorrin-2 C20-
DE   methyltransferase (EC 2.1.1.130) (S-adenosyl-L-methionine--precorrin-2
DE   methyltransferase) (SP2MT); Precorrin-3 methylase (EC 2.1.1.-)].
GN   COBIJ OR COBI OR RV2066 OR MT2126 OR MTCY49.05 OR MB2092.
OS   Mycobacterium tuberculosis, and
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773, 1765;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX   MEDLINE=22206494; PubMed=12218036;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT   "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   J. Bacteriol. 184:5479-5490(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.bovis; STRAIN=AF2122/97;
RX   MEDLINE=22709107; PubMed=12788972;
RA   Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA   Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA   Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA   Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT   "The complete genome sequence of Mycobacterium bovis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC   -!- FUNCTION: METHYLATES PRECORRIN-2 AT THE C-20 POSITION TO PRODUCE
CC       PRECORRIN-3A (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-2 = S-
CC       adenosyl-L-homocysteine + precorrin-3A.
CC   -!- PATHWAY: Cobalamin biosynthesis.
CC   -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS SUMT, CYSG, CBIF/COBM
CC       AND CBIL/COBI.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z73966; CAA98214.1; -.
DR   EMBL; AE007063; AAK46406.1; -.
DR   EMBL; BX248341; CAD96945.1; -.
DR   PIR; E70764; E70764.
DR   TIGR; MT2126; -.
DR   TubercuList; Rv2066; -.
DR   InterPro; IPR006364; CobI_CbiL.
DR   InterPro; IPR006363; CobJ.
DR   InterPro; IPR000878; Cor/por_Metransf.
DR   InterPro; IPR003043; Uropor_Metransf.
DR   Pfam; PF00590; TP_methylase; 2.
DR   TIGRFAMs; TIGR01467; cobI_cbiL; 1.
DR   TIGRFAMs; TIGR01466; cobJ_cbiH; 1.
DR   PROSITE; PS00839; SUMT_1; 1.
DR   PROSITE; PS00840; SUMT_2; 1.
KW   Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW   Methyltransferase; Multifunctional enzyme; Complete proteome.
FT   DOMAIN        1    243       PRECORRIN-2 C20-METHYLTRANSFERASE.
FT   DOMAIN      244    508       PRECORRIN-3 METHYLASE.
SQ   SEQUENCE   508 AA;  53910 MW;  95AC066F022C4DC1 CRC64;

  Query Match           60.0%;  Score 39;  DB 1;  Length 508;
  Best Local Similarity   63.6%;  Pred. No. 46;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 AVVGLSPGEQE 13
              ||||| ||: :
Db        250 AVVGLGPGDSD 260


RESULT 7
NR41_RAT
ID   NR41_RAT       STANDARD;      PRT;   597 AA.
AC   P22829;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Orphan nuclear receptor HMR (Nerve growth factor induced protein I-B)
DE   (NGFI-B) (NUR77).
GN   NR4A1 OR HMR OR NGFIB.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90166506; PubMed=3272167;
RA   Milbrandt J.;
RT   "Nerve growth factor induces a gene homologous to the glucocorticoid
RT   receptor gene.";
RL   Neuron 1:183-188(1988).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=93361012; PubMed=8395013;
RA   Wilson T.E., Fahrner T.J., Milbrandt J.;
RT   "The orphan receptors NGFI-B and steroidogenic factor 1 establish
RT   monomer binding as a third paradigm of nuclear receptor-DNA
RT   interaction.";
RL   Mol. Cell. Biol. 13:5794-5804(1993).
RN   [3]
RP   DNA BINDING MOTIFS.
RX   MEDLINE=92229411; PubMed=1314418;
RA   Wilson T.E., Paulsen R.E., Padgett K.A., Milbrandt J.;
RT   "Participation of non-zinc finger residues in DNA binding by two
RT   nuclear orphan receptors.";
RL   Science 256:107-110(1992).
RN   [4]
RP   PHOSPHORYLATION.
RX   MEDLINE=94043340; PubMed=8227042;
RA   Hirata Y., Kiuchi K., Chen H.-C., Milbrandt J., Guroff G.;
RT   "The phosphorylation and DNA binding of the DNA-binding domain of the
RT   orphan nuclear receptor NGFI-B.";
RL   J. Biol. Chem. 268:24808-24812(1993).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 264-351 IN COMPLEX WITH NBRE,
RP   AND METAL-BINDING.
RX   MEDLINE=99260743; PubMed=10331876;
RA   Meinke G., Sigler P.B.;
RT   "DNA-binding mechanism of the monomeric orphan nuclear receptor
RT   NGFI-B.";
RL   Nat. Struct. Biol. 6:471-477(1999).
CC   -!- FUNCTION: Probable nuclear receptor. May act concomitantly with
CC       NURR1 in regulating the expression of delayed-early genes during
CC       liver regeneration. Binds the NGFI-B response element (NBRE) 5'-
CC       AAAAGGTCA-3'.
CC   -!- SUBUNIT: Binds DNA as a monomer.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Expressed in lung, brain and superior
CC       cervical ganglia. High levels are seen in the adrenal tissue.
CC   -!- INDUCTION: By nerve growth factor and during liver regeneration.
CC   -!- PTM: Phosphorylation of Ser-350 results in decrease in NBRE
CC       binding while phosphorylation of Ser-340 has little effect on it.
CC   -!- SIMILARITY: Belongs to the nuclear hormone receptor family. NR4
CC       subfamily.
CC   ---------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U17254; AAA56770.1; ALT_INIT.
DR   PDB; 1CIT; 26-JUN-00.
DR   TRANSFAC; T00619; -.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR008946; Str_ncl_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Phosphorylation; 3D-structure.
FT   DNA_BIND     266     331       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING      266     286       C4-TYPE.
FT   ZN_FING      302     326       C4-TYPE.
FT   DOMAIN       408     458       LIGAND-BINDING (POTENTIAL).
FT   DOMAIN        80      91       POLY-SER.
FT   DOMAIN       182     186       POLY-PRO.
FT   DOMAIN       582     585       POLY-PRO.
FT   MOD_RES      340     340       PHOSPHORYLATION (BY PKA).
FT   MOD_RES      350     350       PHOSPHORYLATION (BY PKA).
FT   MUTAGEN      340     340       S->A: LOSS OF PHOSPHORYLATION.
FT   MUTAGEN      350     350       S->A: LOSS OF PHOSPHORYLATION.
FT   MUTAGEN      345     345       R->K: DECREASED NBRE BINDING.
FT   MUTAGEN      348     348       L->V: ALMOST COMPLETE LOSS OF NBRE
FT                                  BINDING.
SQ   SEQUENCE   597 AA;  64281 MW;  9CFA987112337E53 CRC64;

  Query Match             60.0%;  Score 39;  DB 1;  Length 597;
  Best Local Similarity   46.2%;  Pred. No. 54;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQE 13
              :|  : ||||:|:
Db        423 IPGFIELSPGDQD 435


RESULT 8
N160_MOUSE
ID   N160_MOUSE     STANDARD;      PRT;  1402 AA.
AC   Q9Z0W3; Q9CZD9;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Nuclear pore complex protein Nup160 (Nucleoporin Nup160) (160 kDa
DE   nucleoporin) (Gene trap locus 1-13) (GTL-13).
GN   NUP160 OR GTL1-13 OR KIAA0197.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SvJ;
RA   Van de Putte T., Cozijnsen M., Dewulf N., Tylzanowski P., Lonnoy O.,
RA   Huylebroeck D.;
RT   "Mus musculus mRNA for gtl-13 (gene trap locus-13), similar to human
RT   KIAA0197 gene (D83781), complete cds.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1151-1402 FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   IDENTIFICATION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX   MEDLINE=21448620; PubMed=11564755;
RA   Belgareh N., Rabut G., Bai S.W., van Overbeek M., Beaudouin J.,
RA   Daigle N., Zatsepina O.V., Pasteau F., Labas V., Fromont-Racine M.,
RA   Ellenberg J., Doye V.;
RT   "An evolutionarily conserved NPC subcomplex, which redistributes in
RT   part to kinetochores in mammalian cells.";
RL   J. Cell Biol. 154:1147-1160(2001).
RN   [4]
RP   IDENTIFICATION, FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX   MEDLINE=21541555; PubMed=11684705;
RA   Vasu S., Shah S., Orjalo A., Park M., Fischer W.H., Forbes D.J.;
RT   "Novel vertebrate nucleoporins Nup133 and Nup160 play a role in mRNA
RT   export.";
RL   J. Cell Biol. 155:339-354(2001).
CC   -!- FUNCTION: Involved in poly(A)+ RNA transport.
CC   -!- SUBUNIT: Forms part of the Nup160 subcomplex in the nuclear pore
CC       which is composed of Nup160, Nup133, Nup107 and Nup96. This
CC       complex plays a role in RNA export and in tethering Nup98 and
CC       Nup153 to the nucleus.
CC   -!- SUBCELLULAR LOCATION: Nuclear pore complex.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC       frameshift in position 1157 and a stop codon in position 1396.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF104415; AAD17922.2; -.
DR   EMBL; AK012715; BAB28429.1; ALT_FRAME.
DR   MGD; MGI:1926227; Nup160.
DR   GO; GO:0005643; C:nuclear pore; IDA.
DR   GO; GO:0005487; F:nucleocytoplasmic transporter activity; IDA.
DR   GO; GO:0006406; P:mRNA-nucleus export; IDA.
KW   Nuclear protein; Transport.
FT   CONFLICT   1156   1156       A -> T (IN REF. 2).
FT   CONFLICT   1314   1314       E -> G (IN REF. 2).
FT   CONFLICT   1368   1368       N -> D (IN REF. 2).
SQ   SEQUENCE   1402 AA;  158230 MW;  3BF5D9F057D28772 CRC64;

  Query Match             60.0%;  Score 39;  DB 1;  Length 1402;
  Best Local Similarity   70.0%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPG 10
```

```
                :|:| |||||
Db        170 IPSVPGLSPG 179


RESULT 9
BUD3_YEAST
ID   BUD3_YEAST      STANDARD;      PRT;  1636 AA.
AC   P25558; P25556; P25557; P87007;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Bud site selection protein BUD3.
GN   BUD3 OR YCL014W/YCL013W/YCL012W OR YCL14W/YCL13W/YCL12W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95247824; PubMed=7730410;
RA   Chant J., Mischke M., Mitchell E., Herskowitz I., Pringle J.R.;
RT   "Role of Bud3p in producing the axial budding pattern of yeast.";
RL   J. Cell Biol. 129:767-778(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=92244356; PubMed=1574125;
RA   Oliver S.G., van der Aart Q.J.M., Agostoni-Carbone M.L., Aigle M.,
RA   Alberghina L., Alexandraki D., Antoine G., Anwar R., Ballesta J.P.G.,
RA   Benit P., Berben G., Bergantino E., Biteau N., Bolle P.-A.,
RA   Bolotin-Fukuhara M., Brown A.J.P., Brown R., Buhler J.-M.,
RA   Carignani G., Chanet R., Contreras R., Crouzet M., Daignan-Fornier B.,
RA   De Haan M., Defoor E., Delgado M.D., Demolder J., Doira C., Dubois E.,
RA   Dujon B., Duesterhoeft A., Erdmann D., Esteban M., Fabre F.,
RA   Fairhead C.A., Faye G., Feldmann H., Fiers W.,
RA   Francingues-Gaillard M.-C., Franco L., Frontali L., Fukuhara H.,
RA   Fuller L.J., Gent M.E., Gigot D., Gilliquet V., Glansdorff N.,
RA   Goffeau A., Grenson M., Grisanti P., Grivell L.A., Haasemann M.,
RA   Hatat D., Hegemann J.H., Herbert C.J., Hilger F., Hohmann S.,
RA   Hollenberg C.P., Huse K., Iborra F., Indge K.J., Isono K., Jackman P.,
RA   Jacq C., Jacquet M., James C.M., Jauniaux J.-C., Jia Y., Jimenez A.,
RA   Kleinhans U., Kreisl P., Lafranchi G., Lewis C., van der Linden C.G.,
RA   Lucchini G., Lutzenkirchen K., Maat C., Mannhaupt G., Manzano M.E.,
RA   Martegani E., Mathieu A., Maurer C.T.C., McConnell D., McKee R.A.,
RA   Messenguy F., Mewes H.-W., Molemans F., Montague M.A., Navas L.,
RA   Newlon C.S., Olson M.V., Pallier C., Panzeri L., Pearson B.M.,
RA   Perea J., Philippsen P., Pierard A., Planta R.J., Plevani P.,
RA   Poetsch B., Pohl F.M., Purnelle B., Ramezani Rad M., Rasmussen S.W.,
RA   Raynal A., Remacha M., Richterich P., Roberts A.B., Rodriguez F.,
RA   Sanz E., Schaaff-Gerstenschlaeger I., Scherens B., Schweitzer B.,
RA   Shu Y., Skala J., Slonimski P.P., Sor F., Soustelle C.,
RA   Spiegelberg R., Stateva L.I., Steensma H.Y., Steiner S., Thierry A.,
RA   Thireos G., Triano L.N., Urrestarazu L.A., Valle G., Vetter I.,
RA   van Vliet-Reedijk J.C., Volckaert G., Vreken P., Warmington J.R.,
RA   von Wettstein D., Wicksteed B.L., Wilson C., Wurst H., Xu G.,
RA   Zimmermann F.K., Sgouros J.G.;
RT   "The complete DNA sequence of yeast chromosome III.";
RL   Nature 357:38-46(1992).
RN   [3]
RP   REVISIONS.
RA   Gromadka R.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   REVISIONS.
RA   Valles G., Volckaerts G.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Co-assembles with BUD4 at bud sites. BUD4 and BUD3 may
CC       cooperate to recognize a spatial landmark (the neck filaments)
CC       during mitosis and they subsequently become a landmark for
CC       establishing the axial budding pattern in G1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17580; AAA86315.1; -.
DR   EMBL; X59720; CAA42346.2; -.
DR   PIR; S74285; S74285.
DR   PIR; S74286; S74286.
DR   GermOnline; 138855; -.
DR   SGD; S0000520; BUD3.
DR   InterPro; IPR000219; RhoGEF.
DR   SMART; SM00325; RhoGEF; 1.
KW   Cell cycle.
SQ   SEQUENCE   1636 AA;  184717 MW;  9E4E46BA5C3A3F69 CRC64;

  Query Match             60.0%;  Score 39;  DB 1;  Length 1636;
  Best Local Similarity   58.3%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 PAVVGLSPGEQE 13
              ||:  ||| :||
Db        112 PAIENLSPSDQE 123


RESULT 10
COBL_MYCTU
ID   COBL_MYCTU      STANDARD;      PRT;   390 AA.
AC   Q10671;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Precorrin-6Y C5,15-methyltransferase [decarboxylating] (EC 2.1.1.132)
DE   (Precorrin-6 methyltransferase) (Precorrin-6Y methylase).
GN   COBL OR RV2072C OR MT2132 OR MTCY49.11C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RX   MEDLINE=22206494; PubMed=12218036;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT   "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   J. Bacteriol. 184:5479-5490(2002).
CC   -!- FUNCTION: CATALYZES THE METHYLATION OF BOTH C-5 AND C-15 IN
CC       PRECORRIN-6Y TO FORM PRECORRIN-8X.
CC   -!- CATALYTIC ACTIVITY: 2 S-adenosyl-L-methionine + precorrin-6Y = 2
CC       S-adenosyl-L-homocysteine + precorrin-8X + CO(2).
CC   -!- PATHWAY: Cobalamin biosynthesis.
CC   -!- SIMILARITY: TO S.TYPHIMURIUM CBIE; ALSO, LOW, TO OTHER
```

```
CC       METHYLASES INVOLVED IN COBALAMIN BIOSYNTHESIS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z73966; CAA98225.1; -.
DR   EMBL; AE007063; AAK46412.1; -.
DR   PIR; C70765; C70765.
DR   TIGR; MT2132; -.
DR   TubercuList; Rv2072c; -.
DR   InterPro; IPR006365; CobL.
DR   InterPro; IPR000878; Cor/por_Metransf.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF00590; TP_methylase; 1.
DR   TIGRFAMs; TIGR01468; cobL_cbiET; 1.
KW   Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW   Methyltransferase; Complete proteome.
FT   CONFLICT    205    205       L -> P (IN REF. 2).
FT   CONFLICT    327    327       D -> H (IN REF. 2).
SQ   SEQUENCE   390 AA;  41854 MW;  FB42EFF7562F21F3 CRC64;

  Query Match             58.5%;  Score 38;  DB 1;  Length 390;
  Best Local Similarity   88.9%;  Pred. No. 53;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSP 9
              |||| ||||
Db         55 LPAVQGLSP 63


RESULT 11
RS1_LEULA
ID   RS1_LEULA      STANDARD;      PRT;   429 AA.
AC   P50889; P71450;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   40S ribosomal protein S1.
GN   RPS1.
OS   Leuconostoc lactis.
OC   Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX   NCBI_TaxID=1246;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97186703; PubMed=9034319;
RA   Yamit-Hezi A., Levy Z., Neuman S., Nudel U.;
RT   "A Leuconostoc lactis protein with homology to ribosomal protein S1
RT   shares common epitopes and common DNA binding properties with a
RT   mammalian DNA binding nuclear factor.";
RL   Gene 185:99-103(1997).
RN   [2]
RP   SEQUENCE OF 24-429 FROM N.A.
RX   MEDLINE=95237615; PubMed=7721096;
RA   Eklund E.A., Lee S.W., Skalnik D.G.;
RT   "Cloning of a cDNA encoding a human DNA-binding protein similar to
RT   ribosomal protein S1.";
RL   Gene 155:231-235(1995).
RN   [3]
RP   SEQUENCE OF 78-429 FROM N.A.
RX   MEDLINE=96164600; PubMed=8568274;
RA   Tsuzaka K., Leu A.K., Frank M.B., Movafagh B.F., Koscec M.,
RA   Winkler T.H., Kalden J.R., Reichlin M.;
RT   "Lupus autoantibodies to double-stranded DNA cross-react with
RT   ribosomal protein S1.";
RL   J. Immunol. 156:1668-1675(1996).
CC   -!- FUNCTION: EXHIBITS PREFERENTIAL BINDING TO SINGLE-STRANDED AND
CC       DOUBLE-STRANDED DNA AND A LOW BINDING AFFINITY FOR RNA.
CC   -!- SIMILARITY: Belongs to the S1P family of ribosomal proteins.
CC   -!- SIMILARITY: Contains 4 S1 motif domains.
CC   -!- CAUTION: WAS ORIGINALLY (REF.2 AND REF.3) THOUGHT TO ORIGINATE
CC       FROM HUMAN BUT IS MOST PROBABLY THE RESULT OF A CDNA LIBRARY
CC       CONTAMINATION BY L.LACTIS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U24086; AAB08978.1; -.
DR   EMBL; U05589; AAA77669.1; -.
DR   EMBL; U27517; AAA97575.1; -.
DR   HSSP; P05055; 1SRO.
DR   InterPro; IPR008994; Nucleic_acid_OB.
DR   InterPro; IPR000110; Ribosomal_S1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF00575; S1; 4.
DR   PRINTS; PR00681; RIBOSOMALS1.
DR   SMART; SM00316; S1; 4.
DR   PROSITE; PS50126; S1; 4.
KW   Ribosomal protein; Repeat; RNA-binding.
FT   DOMAIN       55    128       S1 MOTIF 1.
FT   DOMAIN      144    211       S1 MOTIF 2.
FT   DOMAIN      231    299       S1 MOTIF 3.
FT   DOMAIN      316    385       S1 MOTIF 4.
FT   CONFLICT     24     24       S -> G (IN REF. 2).
FT   CONFLICT    122    122       A -> S (IN REF. 3).
FT   CONFLICT    217    217       L -> R (IN REF. 2 AND 3).
SQ   SEQUENCE   429 AA;  46386 MW;  92AC82605F39DDFC CRC64;

  Query Match             58.5%;  Score 38;  DB 1;  Length 429;
  Best Local Similarity   80.0%;  Pred. No. 58;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AVVGLSPGEQ 12
              |||||| ||:
Db         71 AVVGLSTGEE 80


RESULT 12
CATA_PICAN
ID   CATA_PICAN     STANDARD;      PRT;   507 AA.
AC   P30263;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Peroxisomal catalase (EC 1.11.1.6).
GN   PXP9 OR PXP-9.
OS   Pichia angusta (Yeast) (Hansenula polymorpha).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Pichia.
OX   NCBI_TaxID=4905;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 34438;
RX   MEDLINE=92299073; PubMed=1607006;
RA   Didion T., Roggenkamp R.O.;
RT   "Targeting signal of the peroxisomal catalase in the methylotrophic
RT   yeast Hansenula polymorpha.";
RL   FEBS Lett. 303:113-116(1992).
CC   -!- FUNCTION: Occurs in almost all aerobically respiring organisms and
CC       serves to protect cells from the toxic effects of hydrogen
CC       peroxide.
CC   -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC   -!- COFACTOR: Heme group.
CC   -!- SUBUNIT: Homotetramer.
CC   -!- SUBCELLULAR LOCATION: Peroxisomal.
CC   -!- SIMILARITY: Belongs to the catalase family.
CC   ---------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X56501; CAA39856.1; -.
DR   PIR; S23422; S23422.
DR   HSSP; P15202; 1A4E.
DR   InterPro; IPR002226; Catalase.
DR   Pfam; PF00199; catalase; 1.
DR   PRINTS; PR00067; CATALASE.
DR   ProDom; PD000510; Catalase; 1.
DR   PROSITE; PS00437; CATALASE_1; 1.
DR   PROSITE; PS00438; CATALASE_2; 1.
KW   Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
KW   Peroxisome.
FT   ACT_SITE     65     65       BY SIMILARITY.
FT   ACT_SITE    138    138       BY SIMILARITY.
FT   METAL       348    348       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT   SITE        505    507       MICROBODY TARGETING SIGNAL (POTENTIAL).
SQ   SEQUENCE   507 AA;  57849 MW;  3536ED0A49539CC3 CRC64;

  Query Match             58.5%;  Score 38;  DB 1;  Length 507;
  Best Local Similarity   70.0%;  Pred. No. 68;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 VVGLSPGEQE 13
              |:| :|||||
Db        445 VLGRTPGEQE 454


RESULT 13
C4AD_DROME
ID   C4AD_DROME     STANDARD;      PRT;   516 AA.
AC   Q9V4T3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable cytochrome P450 4ad1 (EC 1.14.-.-) (CYPIVAD1).
GN   CYP4AD1 OR CG2110.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley; TISSUE=Embryo;
RX   MEDLINE=22426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC   -!- FUNCTION: May be involved in the metabolism of insect hormones and
CC       in the breakdown of synthetic insecticides (By similarity).
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum
CC       (Potential).
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE003837; AAF59092.1; -.
DR   EMBL; AY061058; AAL28606.1; -.
DR   HSSP; P14779; 1JPZ.
DR   FlyBase; FBgn0033292; Cyp4ad1.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW   Endoplasmic reticulum; Hypothetical protein.
FT   METAL       445    445       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   516 AA;  58870 MW;  648EA22492AF58C7 CRC64;

  Query Match             58.5%;  Score 38;  DB 1;  Length 516;
  Best Local Similarity   80.0%;  Pred. No. 69;
  Matches    8;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPG 10
              |||| || ||
Db        469 LPAVDGLPPG 478


RESULT 14
EFG_HAEIN
ID   EFG_HAEIN      STANDARD;      PRT;   699 AA.
AC   P43925;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Elongation factor G (EF-G).
GN   FUSA OR FUS OR HI0579.
```

```
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: This protein promotes the GTP-dependent translocation of
CC       the nascent protein chain from the A-site to the P-site of the
CC       ribosome.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the GTP-binding elongation factor family.
CC       EF-G/EF-2 subfamily.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U32739; AAC22237.1; -.
DR   PIR; F64078; F64078.
DR   HSSP; P13551; 1ELO.
DR   TIGR; HI0579; -.
DR   HAMAP; MF_00054; -; 1.
DR   InterPro; IPR004540; EF-G.
DR   InterPro; IPR000795; EF_GTPbind.
DR   InterPro; IPR000640; EFG_C.
DR   InterPro; IPR009022; EFG_III_V.
DR   InterPro; IPR005517; EFG_IV.
DR   InterPro; IPR004161; EFTU_D2.
DR   InterPro; IPR005225; Small_GTP.
DR   InterPro; IPR009000; Translat_factor.
DR   Pfam; PF00679; EFG_C; 1.
DR   Pfam; PF03764; EFG_IV; 1.
DR   Pfam; PF00009; GTP_EFTU; 1.
DR   Pfam; PF03144; GTP_EFTU_D2; 1.
DR   PRINTS; PR00315; ELONGATNFCT.
DR   TIGRFAMs; TIGR00484; EF-G; 1.
DR   TIGRFAMs; TIGR00231; small_GTP; 1.
DR   PROSITE; PS00301; EFACTOR_GTP; 1.
KW   Elongation factor; Protein biosynthesis; GTP-binding;
KW   Complete proteome.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   NP_BIND      16     23       GTP (BY SIMILARITY).
FT   NP_BIND      87     91       GTP (BY SIMILARITY).
FT   NP_BIND     141    144       GTP (BY SIMILARITY).
SQ   SEQUENCE   699 AA;  77132 MW;  FBBADE39C0F62801 CRC64;

  Query Match             58.5%;  Score 38;  DB 1;  Length 699;
  Best Local Similarity   46.2%;  Pred. No. 93;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQE 13
              :||: |::| | |
Db        291 IPAIKGINPDETE 303
```

```
RESULT 15
EFG_PASMU
ID   EFG_PASMU      STANDARD;      PRT;   700 AA.
AC   P57938;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Elongation factor G (EF-G).
GN   FUSA OR PM1356.
OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Pasteurella.
OX   NCBI_TaxID=747;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pm70;
RX   MEDLINE=21145866; PubMed=11248100;
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC   -!- FUNCTION: This protein promotes the GTP-dependent translocation of
CC       the nascent protein chain from the A-site to the P-site of the
CC       ribosome.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the GTP-binding elongation factor family.
CC       EF-G/EF-2 subfamily.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE006173; AAK03440.1; -.
DR   HSSP; P13551; 1ELO.
DR   HAMAP; MF_00054; -; 1.
DR   InterPro; IPR004540; EF-G.
DR   InterPro; IPR000795; EF_GTPbind.
DR   InterPro; IPR000640; EFG_C.
DR   InterPro; IPR009022; EFG_III_V.
DR   InterPro; IPR005517; EFG_IV.
DR   InterPro; IPR004161; EFTU_D2.
DR   InterPro; IPR005225; Small_GTP.
DR   InterPro; IPR009000; Translat_factor.
DR   Pfam; PF00679; EFG_C; 1.
DR   Pfam; PF03764; EFG_IV; 1.
DR   Pfam; PF00009; GTP_EFTU; 1.
DR   Pfam; PF03144; GTP_EFTU_D2; 1.
DR   PRINTS; PR00315; ELONGATNFCT.
DR   TIGRFAMs; TIGR00484; EF-G; 1.
DR   TIGRFAMs; TIGR00231; small_GTP; 1.
DR   PROSITE; PS00301; EFACTOR_GTP; 1.
KW   Elongation factor; Protein biosynthesis; GTP-binding;
KW   Complete proteome.
FT   NP_BIND      16     23       GTP (BY SIMILARITY).
FT   NP_BIND      87     91       GTP (BY SIMILARITY).
FT   NP_BIND     141    144       GTP (BY SIMILARITY).
SQ   SEQUENCE   700 AA;  77186 MW;  6CC161F7F9FA9C72 CRC64;

  Query Match             58.5%;  Score 38;  DB 1;  Length 700;
  Best Local Similarity   46.2%;  Pred. No. 93;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQE 13
              :||: |::| | |
Db        292 IPAIKGINPDETE 304


Search completed: May  7, 2004, 12:34:33
Job time : 4.68 secs
```

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:27:20 ; Search time 16 Seconds
(without alignments)
150.299 Million cell updates/sec

Title: US-09-786-214A-5
Perfect score: 130
Sequence: 1 MAGLPAVVGLSPGEQEYHRGGVGVL 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	51	39.2	426	2	D72635	hypothetical prote
2	50.5	38.8	615	2	AC3554	galactonate dehydr
3	50	38.5	277	2	F82130	formyltetrahydrofo
4	50	38.5	381	2	F75270	cytochrome P450 -
5	49.5	38.1	603	2	AE2952	dihydroxy-acid deh
6	49.5	38.1	603	2	H98330	hypothetical prote
7	49	37.7	118	2	S38491	Ig heavy chain - h
8	49	37.7	278	2	E64131	formyltetrahydrofo
9	48.5	37.3	559	2	S62503	inorganic phosphat
10	48	36.9	522	2	T44369	pyruvate, water di
11	48	36.9	567	2	AF0274	ribulokinase (EC 2
12	48	36.9	780	2	D75361	phosphoenolpyruvat
13	48	36.9	792	2	B82752	penicillin binding
14	47	36.2	280	2	F90845	formyltetrahydrofo
15	47	36.2	280	2	E85703	hypothetical prote
16	47	36.2	280	2	C36871	formyltetrahydrofo
17	47	36.2	280	2	AF0649	formyltetrahydrofo
18	47	36.2	293	2	E71819	formyltetrahydrofo
19	47	36.2	293	2	B64699	formyltetrahydrofo
20	47	36.2	661	2	G84511	hypothetical prote
21	47	36.2	2109	1	I50421	aggrecan precursor
22	46.5	35.8	178	2	G82977	hypothetical prote
23	46.5	35.8	563	2	T36580	hypothetical prote
24	46.5	35.8	656	2	B81692	penicillin-binding
25	46	35.4	192	2	AB2545	hypothetical prote
26	46	35.4	271	2	C82841	cysteine proteinas
27	46	35.4	274	2	F81350	formyltetrahydrofo
28	46	35.4	287	2	B70871	hypothetical prote
29	46	35.4	300	2	C69857	formyltetrahydrofo
30	46	35.4	508	2	E70764	probable cobI prot
31	46	35.4	638	2	F75547	anthranilate synth
32	46	35.4	1196	2	H87542	hydantoinase/oxopr
33	45.5	35.0	717	2	T50685	delta1-pyrroline-5
34	45	34.6	104	2	C72637	hypothetical prote
35	45	34.6	158	2	AI2718	conserved hypothet
36	45	34.6	158	2	E97500	hypothetical prote
37	45	34.6	211	2	AE3572	carbonate dehydrat
38	45	34.6	217	2	E72047	conserved hypothet
39	45	34.6	217	2	D86576	CT482 hypothetical
40	45	34.6	340	2	AI2075	transcription regu
41	45	34.6	463	2	T49460	hypothetical prote
42	45	34.6	479	2	F82432	PTS system, sucros
43	45	34.6	479	2	JQ0781	sucrose uptake pro
44	45	34.6	582	2	T34538	hypothetical prote
45	45	34.6	940	2	JE0291	FB19 protein - hum

ALIGNMENTS

RESULT 1
D72635
hypothetical protein APE1539 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
C;Accession: D72635
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropy
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: D72635
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-426 <KAW>
A;Cross-references: DDBJ:AP000061; NID:g5104821; PIDN:BAA80538.1; PID:d1044324; PID:g51
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1539
C;Superfamily: Aeropyrum pernix hypothetical protein APE1539

Query Match 39.2%; Score 51; DB 2; Length 426;
Best Local Similarity 45.5%; Pred. No. 15;
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

```
Qy          3 GLPAVVGLSPGEQEYHRGGVGV 24
              ||  ::|: || :|  | || |
Db        278 GLADILGIEPGRREERREGVRV 299
```

RESULT 2
AC3554
galactonate dehydratase (EC 4.2.1.6) [imported] - Brucella melitensis (strain 16M)
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C;Accession: AC3554
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, .; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letes: Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella meliten:
A;Reference number: AD3252; PMID:11756688
A;Accession: AC3554
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-615 <KUR>
A;Cross-references: GB:AE008918; PIDN:AAL53598.1; PID:g17984511; GSPDB:GN00191
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEII0356
A;Map position: II
C;Superfamily: dihydroxy-acid dehydratase
C;Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 38.8%; Score 50.5; DB 2; Length 615;
Best Local Similarity 45.8%; Pred. No. 26;
Matches 11; Conservative 2; Mismatches 6; Indels 5; Gaps 1;

```
Qy          4 LPAVVGLSP-----GEQEYHRGGV 22
              :| :| | |       || || |||
Db        324 IPLLVNLQPAGEYLGEDYYHAGGV 347
```

RESULT 3
F82130
formyltetrahydrofolate deformylase VC1992 [imported] - Vibrio cholerae (strain N16961 se
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: F82130
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: F82130
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-277 <HEI>
A;Cross-references: GB:AE004274; GB:AE003852; NID:g9656533; PIDN:AAF95140.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC1992
A;Map position: 1
C;Superfamily: phosphoribosylglycinamide formyltransferase; phosphoribosylglycinamide fo

Query Match 38.5%; Score 50; DB 2; Length 277;
Best Local Similarity 52.9%; Pred. No. 13;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

```
Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :|  | :| | ||
Db        190 LPAFIGAKPYQQAYERG 206
```

RESULT 4
F75270
cytochrome P450 - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: F75270
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: F75270
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-381 <WHI>
A;Cross-references: GB:AE002076; GB:AE000513; NID:g6460285; PIDN:AAF12016.1; PID:g646029
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2473
A;Map position: 1

Query Match 38.5%; Score 50; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          3 GLPAVVGLSP 12
              ||||||||||
Db         51 GLPAVVGLSP 60
```

RESULT 5
AE2952
dihydroxy-acid dehydratase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AE2952
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClel
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AE2952
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-603 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44035.1; PID:g17741597; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: ilvD
A;Map position: linear chromosome
C;Superfamily: dihydroxy-acid dehydratase

Query Match 38.1%; Score 49.5; DB 2; Length 603;
Best Local Similarity 45.8%; Pred. No. 36;
Matches 11; Conservative 2; Mismatches 6; Indels 5; Gaps 1;

```
Qy          4 LPAVVGLSP-----GEQEYHRGGV 22
              :| :| | |       || || |||
Db        312 VPLLVNLQPAGEYLGEDYYHAGGV 335
```

RESULT 6
H98330
hypothetical protein AGR_L_3190 [imported] - Agrobacterium tumefaciens (strain C58, Cer
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: H98330
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tu
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: H98330
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-603 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK90170.1; PID:g15160173; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_3190
A;Map position: linear chromosome
C;Superfamily: dihydroxy-acid dehydratase

Query Match 38.1%; Score 49.5; DB 2; Length 603;
Best Local Similarity 45.8%; Pred. No. 36;
Matches 11; Conservative 2; Mismatches 6; Indels 5; Gaps 1;

```
Qy          4 LPAVVGLSP-----GEQEYHRGGV 22
              :| :| | |       || || |||
Db        312 VPLLVNLQPAGEYLGEDYYHAGGV 335
```

RESULT 7
S38491
Ig heavy chain - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 23-Jul-1999
C;Accession: S38491
R;Marks, J.D.; Ouwehand, W.H.; Bye, J.M.; Finnern, R.; Gorick, B.D.; Voak, D.; Thorpe,
submitted to the EMBL Data Library, June 1993
A;Description: Human antibody fragments specific for human blood group antigens from a

A;Reference number: S38488
A;Accession: S38491
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-118 <MAR>
A;Cross-references: EMBL:Z23032; NID:g414029; PIDN:CAA80567.1; PID:g414030
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin
F;15-97/Domain: immunoglobulin homology <IMM>

Query Match 37.7%; Score 49; DB 2; Length 118;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Qy 3 GLPAVVGLSPGEQEYHRGGV 22
Db 44 GLEWVSGIGPGGDTYHPGSV 63

RESULT 8
E64131
formyltetrahydrofolate deformylase (EC 3.5.1.10) - Haemophilus influenzae (strain Rd KW2
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 17-Mar-2000
C;Accession: E64131; PN0606
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: E64131
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-278 <TIGR>
A;Cross-references: GB:U32833; GB:L42023; NID:g1574432; PIDN:AAC23236.1; PID:g1574433; T
R;Maskell, D.
Gene 129, 155-156, 1993
A;Title: Cloning and sequencing of the Haemophilus influenzae aroA gene.
A;Reference number: JN0758; MUID:93328119; PMID:8335255
A;Accession: PN0606
A;Molecule type: DNA
A;Residues: 64-114,'RNR',118-137,'PK',141-204,'E',206-278 <MAS>
A;Cross-references: GB:L04686; NID:g148863; PIDN:AAA24942.1; PID:g148864
C;Genetics:
A;Gene: purN
C;Function:
A;Description: catalyzes hydrolysis of 10-formyltetrahydrofolate to tetrahydrofolate and
A;Pathway: one-carbon metabolism
A;Note: activated by methionine and inhibited by glycine
C;Superfamily: phosphoribosylglycinamide formyltransferase; phosphoribosylglycinamide fo
C;Keywords: hydrolase
F;85-277/Domain: phosphoribosylglycinamide formyltransferase homology <PRGF>
F;223/Active site: Asp #status predicted

Query Match 37.7%; Score 49; DB 2; Length 278;
Best Local Similarity 52.9%; Pred. No. 18;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 4 LPAVVGLSPGEQEYHRG 20
Db 191 LPAFIGAKPYQQAYKRG 207

RESULT 9
S62503
inorganic phosphate transporter - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 10-Dec-1999
C;Accession: T38287; S62503
R;Niblett, D.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, October 1995
A;Reference number: Z21783
A;Accession: T38287
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-559 <NI2>
A;Cross-references: EMBL:Z64354; NID:g1039338; PIDN:CAA91247.1; PID:g1039350; GSPDB:GN0
A;Experimental source: strain 972h-; cosmid c23D3
C;Genetics:
A;Gene: SPDB:SPAC23D3.12
A;Map position: 1R
C;Superfamily: probable inorganic phosphate transport protein PHO84

Query Match 37.3%; Score 48.5; DB 2; Length 559;
Best Local Similarity 47.6%; Pred. No. 46;
Matches 10; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

Qy 8 VGLSPGEQEYH---RGGVGVL 25
Db 365 IGFSSGKNEYHTLMRGAIGNL 385

RESULT 10
T44369
pyruvate, water dikinase (EC 2.7.9.2) [imported] - Deinococcus radiodurans (fragment)
C;Species: Deinococcus radiodurans
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 03-Jun-2002
C;Accession: T44369
R;Narumi, I.; Islam, S.; Cherdchu, K.; Kikuchi, M.; Watanabe, H.; Kitayama, S.; Yamamot
submitted to the EMBL Data Library, August 1998
A;Description: IS8301: the second insertion sequence element from Deinococcus radiodura
A;Reference number: Z22755
A;Accession: T44369
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-522 <NAR>
A;Cross-references: EMBL:AB016803; PIDN:BAA32387.1
A;Experimental source: strain KD8301
C;Genetics:
A;Note: ppsA
C;Superfamily: Escherichia coli pyruvate,water dikinase; phosphotransferase system enzy
C;Keywords: transferase

Query Match 36.9%; Score 48; DB 2; Length 522;
Best Local Similarity 50.0%; Pred. No. 50;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 3 GLPAVVGLSPGEQEYHRG 20
Db 160 GIPAVVGTGNATRELHNG 177

RESULT 11
AF0274
ribulokinase (EC 2.7.1.16) [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C;Accession: AF0274
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AF0274
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-567 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC91058.1; PID:g15980250; GSPDB:GN00175
C;Genetics:
A;Gene: araB
C;Superfamily: ribulokinase
C;Keywords: phosphotransferase

```
  Query Match             36.9%;  Score 48;  DB 2;  Length 567;
  Best Local Similarity   52.4%;  Pred. No. 55;
  Matches   11;  Conservative    1;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          3 GLPAVVGLSPGEQEYHRGGVG 23
              |||  | || |  : | | ||
Db        270 GLPTTVILSGGAFDCHMGAVG 290


RESULT 12
D75361
phosphoenolpyruvate synthase - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: D75361
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: D75361
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-780 <WHI>
A;Cross-references: GB:AE002014; GB:AE000513; NID:g6459494; PIDN:AAF11283.1; PID:g645949
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR1727
A;Map position: 1
C;Superfamily: Escherichia coli pyruvate,water dikinase; phosphotransferase system enzym

  Query Match             36.9%;  Score 48;  DB 2;  Length 780;
  Best Local Similarity   50.0%;  Pred. No. 77;
  Matches    9;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          3 GLPAVVGLSPGEQEYHRG 20
              |:|||||      :| | |
Db        418 GIPAVVGTGNATRELHNG 435


RESULT 13
B82752
penicillin binding protein 1B XF0884 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: B82752
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: B82752
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-792 <SIM>
A;Cross-references: GB:AE003927; GB:AE003849; NID:g9105783; PIDN:AAF83694.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF0884

  Query Match             36.9%;  Score 48;  DB 2;  Length 792;
  Best Local Similarity   42.9%;  Pred. No. 78;
  Matches    9;  Conservative    3;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          2 AGLPAVVGLSPGEQEYHRGGV 22
              ||:  : |:||  | |   ||
Db        385 AGMTVLTGMSPSAQSYAERGV 405


RESULT 14
F90845
formyltetrahydrofolate deformylase [imported] - Escherichia coli (strain O157:H7, subst
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: F90845
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: F90845
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-280 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB35157.1; PID:g13361199; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs1734
C;Superfamily: phosphoribosylglycinamide formyltransferase; phosphoribosylglycinamide f

  Query Match             36.2%;  Score 47;  DB 2;  Length 280;
  Best Local Similarity   52.9%;  Pred. No. 36;
  Matches    9;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :|  |  | | ||
Db        193 LPAFIGARPYHQAYERG 209


RESULT 15
E85703
hypothetical protein purU [imported] - Escherichia coli (strain O157:H7, substrain EDL9
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: E85703
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: E85703
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-280 <STO>
A;Cross-references: GB:AE005174; NID:g12514950; PIDN:AAG56089.1; GSPDB:GN00145; UWGP:Z2
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: purU
C;Superfamily: phosphoribosylglycinamide formyltransferase; phosphoribosylglycinamide f

  Query Match             36.2%;  Score 47;  DB 2;  Length 280;
  Best Local Similarity   52.9%;  Pred. No. 36;
  Matches    9;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :|  |  | | ||
Db        193 LPAFIGARPYHQAYERG 209


Search completed: May 7, 2004, 12:39:05
Job time : 17.1667 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:20:30 ; Search time 9 Seconds
(without alignments)
144.639 Million cell updates/sec

Title: US-09-786-214A-5
Perfect score: 130
Sequence: 1 MAGLPAVVGLSPGEQEYHRGGVGVL 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	49	37.7	278	1	PURU_HAEIN	Q03432 haemophilus
2	48.5	37.3	559	1	YAEC_SCHPO	Q09852 schizosacch
3	48	36.9	563	1	ARAB_YERPE	P58543 yersinia pe
4	48	36.9	780	1	PPSA_DEIRA	O83026 deinococcus
5	47	36.2	280	1	PURU_ECOL6	P38480 escherichia
6	47	36.2	280	1	PURU_ECOLI	P37051 escherichia
7	47	36.2	567	1	ARAB_VIBPA	Q87fk5 vibrio para
8	47	36.2	813	1	CADM_MOUSE	Q9wtp5 mus musculu
9	47	36.2	2109	1	PGCA_CHICK	P07898 gallus gall
10	46	35.4	508	1	COBI_MYCTU	Q10677 mycobacteri
11	46	35.4	1172	1	CNA2_MOUSE	Q80xc6 mus musculu
12	45	34.6	479	1	PTSB_VIBAL	P22825 vibrio algi
13	45	34.6	961	1	ACON_MYCAV	O08451 mycobacteri
14	44.5	34.2	618	1	ILVD_CHRVO	Q7nyj7 chromobacte
15	44.5	34.2	2054	1	M18A_HUMAN	Q92614 homo sapien
16	44	33.8	151	1	SODC_XIPGL	P03946 xiphias gla
17	44	33.8	433	1	ENOB_RABIT	P25704 oryctolagus
18	44	33.8	565	1	ARAB_ECO57	P58541 escherichia
19	44	33.8	565	1	ARAB_ECOL6	Q8fl88 escherichia
20	44	33.8	565	1	ARAB_ECOLI	P08204 escherichia
21	44	33.8	568	1	ARAB_SALTI	P58542 salmonella
22	44	33.8	614	1	CPR1_DROME	Q9vn93 drosophila
23	44	33.8	617	1	ILVD_STRCO	O69198 streptomyce
24	44	33.8	658	1	VG18_BPT4	P13332 bacteriopha
25	44	33.8	677	1	SG1_HUMAN	P05060 homo sapien
26	44	33.8	777	1	L100_ADE41	P11824 human adeno
27	44	33.8	813	1	CADM_RAT	Q63315 rattus norv
28	44	33.8	1065	1	RPOB_MARPO	P06272 marchantia
29	44	33.8	3664	1	MINT_HUMAN	Q96t58 homo sapien
30	44	33.8	4351	1	FAT2_RAT	O88277 rattus norv
31	43.5	33.5	173	1	RUVC_XYLFA	Q9pc76 xylella fas
32	43.5	33.5	173	1	RUVC_XYLFT	Q87d03 xylella fas
33	43.5	33.5	968	1	CTD1_HUMAN	O60716 homo sapien
34	43	33.1	211	1	COBL_METJA	Q58917 methanococc
35	43	33.1	325	1	CA19_RAT	P20850 rattus norv
36	43	33.1	326	1	MER_METTI	Q9uxp0 methanolobu
37	43	33.1	364	1	COAI_SV40	P03087 simian viru
38	43	33.1	382	1	FETB_HUMAN	Q9ugm5 homo sapien
39	43	33.1	390	1	COBL_MYCTU	Q10671 mycobacteri
40	43	33.1	481	1	GLGA_RHILO	Q985p2 rhizobium l
41	43	33.1	505	1	YT6J_CAEEL	P90838 caenorhabdi
42	43	33.1	516	1	C4AD_DROME	Q9v4t3 drosophila
43	43	33.1	544	1	CH60_PROAC	Q9k2u4 propionibac
44	43	33.1	561	1	LCFA_SALTY	Q8xgg8 salmonella
45	43	33.1	628	1	HNFA_MOUSE	P22361 mus musculu

ALIGNMENTS

```
RESULT 1
PURU_HAEIN
ID   PURU_HAEIN     STANDARD;      PRT;   278 AA.
AC   Q03432;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-FH(4)
DE   hydrolase).
GN   PURU OR HI1588.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
RN   [2]
RP   SEQUENCE OF 64-278 FROM N.A.
RC   STRAIN=RM 7004 / Serotype B;
RX   MEDLINE=93328119; PubMed=8335255;
RA   Maskell D.J.;
RT   "Cloning and sequencing of the Haemophilus influenzae aroA gene.";
RL   Gene 129:155-156(1993).
CC   -!- FUNCTION: PRODUCES FORMATE FROM FORMYL-TETRAHYDROFOLATE. PROVIDES
CC       THE MAJOR SOURCE OF FORMATE FOR THE PURT-DEPENDENT SYNTHESIS OF
CC       5'-PHOSPHORIBOSYL-N-FORMYLGLYCINAMIDE (FGAR) DURING AEROBIC
CC       GROWTH. HAS A ROLE IN REGULATING THE ONE-CARBON POOL
CC       (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + H(2)O = formate +
CC       tetrahydrofolate.
CC   -!- ENZYME REGULATION: Activated by methionine, inhibited by glycine
CC       (By similarity).
CC   -!- PATHWAY: De novo purine biosynthesis.
CC   -!- SUBUNIT: Homohexamer (By similarity).
CC   -!- SIMILARITY: SOME, TO GAR TRANSFORMYLASE (PURN).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; U32833; AAC23236.1; -.
DR   EMBL; L04686; AAA24942.1; -.
DR   PIR; E64131; E64131.
DR   HSSP; P08179; 1GRC.
DR   TIGR; HI1588; -.
DR   InterPro; IPR002912; ACT.
DR   InterPro; IPR002376; formyl_transf.
DR   InterPro; IPR004810; PurU.
DR   Pfam; PF01842; ACT; 1.
DR   Pfam; PF00551; formyl_transf; 1.
DR   PRINTS; PR01575; FFH4HYDRLASE.
DR   TIGRFAMs; TIGR00655; PurU; 1.
KW   Purine biosynthesis; Hydrolase; One-carbon metabolism;
KW   Complete proteome.
FT   ACT_SITE    223    223       BY SIMILARITY.
FT   CONFLICT    115    117       VIG -> RNR (IN REF. 2).
FT   CONFLICT    138    140       HEN -> PK (IN REF. 2).
FT   CONFLICT    205    205       K -> E (IN REF. 2).
SQ   SEQUENCE   278 AA;  32173 MW;  7F375AB3C422EC4B CRC64;

  Query Match             37.7%;  Score 49;  DB 1;  Length 278;
  Best Local Similarity   52.9%;  Pred. No. 8.6;
  Matches    9;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :|  | :| | ||
Db        191 LPAFIGAKPYQQAYKRG 207


RESULT 2
YAEC_SCHPO
ID   YAEC_SCHPO     STANDARD;      PRT;   559 AA.
AC   Q09852;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Putative inorganic phosphate transporter C23D3.12.
GN   SPAC23D3.12.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: High-affinity transporter for external inorganic
CC       phosphate (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. STRONG, TO
CC       YEAST PHO84.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; Z64354; CAA91247.1; -.
DR   PIR; T38287; S62503.
DR   GeneDB_SPombe; SPAC23D3.12; -.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR004738; Phos_permease.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   TIGRFAMs; TIGR00887; 2A0109; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Hypothetical protein; Phosphate transport; Transport; Transmembrane.
FT   DOMAIN        1     45       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     46     66       1 (POTENTIAL).
FT   DOMAIN       67     94       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     95    115       2 (POTENTIAL).
FT   DOMAIN      116    118       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    119    139       3 (POTENTIAL).
FT   DOMAIN      140    144       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    145    165       4 (POTENTIAL).
FT   DOMAIN      166    188       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    189    209       5 (POTENTIAL).
FT   DOMAIN      210    224       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    225    245       6 (POTENTIAL).
FT   DOMAIN      246    335       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    336    356       7 (POTENTIAL).
FT   DOMAIN      357    381       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    382    402       8 (POTENTIAL).
FT   DOMAIN      403    409       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    410    430       9 (POTENTIAL).
FT   DOMAIN      431    438       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    439    459       10 (POTENTIAL).
FT   DOMAIN      460    474       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    475    495       11 (POTENTIAL).
FT   DOMAIN      496    497       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    498    518       12 (POTENTIAL).
FT   DOMAIN      519    559       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   559 AA;  62316 MW;  2E8958F86C2092E2 CRC64;

  Query Match             37.3%;  Score 48.5;  DB 1;  Length 559;
  Best Local Similarity   47.6%;  Pred. No. 21;
  Matches   10;  Conservative    3;  Mismatches    5;  Indels    3;  Gaps    1;

Qy          8 VGLSPGEQEYH---RGGVGVL 25
              :| | |: |||   || :| |
Db        365 IGFSSGKNEYHTLMRGAIGNL 385


RESULT 3
ARAB_YERPE
ID   ARAB_YERPE     STANDARD;      PRT;   563 AA.
AC   P58543;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   L-ribulokinase (EC 2.7.1.16).
```

```
GN   ARAB OR YPO2254 OR Y2095.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / Biovar Orientalis;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM5 / Biovar Mediaevalis;
RX   MEDLINE=22137863; PubMed=12142430;
RA   Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA   Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA   Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA   Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA   Perry R.D.;
RT   "Genome sequence of Yersinia pestis KIM.";
RL   J. Bacteriol. 184:4601-4611(2002).
CC   -!- CATALYTIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-
CC       phosphate.
CC   -!- PATHWAY: L-arabinose catabolism; second step.
CC   -!- SIMILARITY: Belongs to the ribulokinase family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AJ414151; CAC91058.1; ALT_INIT.
DR   EMBL; AE013812; AAM85658.1; ALT_INIT.
DR   HAMAP; MF_00520; -; 1.
DR   InterPro; IPR000577; FGGY_kin.
DR   InterPro; IPR005929; L_ribulokin.
DR   Pfam; PF00370; FGGY; 1.
DR   Pfam; PF02782; FGGY_C; 1.
DR   TIGRFAMs; TIGR01234; L-ribulokinase; 1.
KW   Transferase; Kinase; Arabinose catabolism; Complete proteome.
SQ   SEQUENCE   563 AA;  61833 MW;  88EF1E0B00A5414E CRC64;

  Query Match             36.9%;  Score 48;  DB 1;  Length 563;
  Best Local Similarity   52.4%;  Pred. No. 25;
  Matches   11;  Conservative    1;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          3 GLPAVVGLSPGEQEYHRGGVG 23
              |||  | || |  : | | ||
Db        266 GLPTTVILSGGAFDCHMGAVG 286


RESULT 4
PPSA_DEIRA
ID   PPSA_DEIRA     STANDARD;      PRT;   780 AA.
AC   Q83026;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phosphoenolpyruvate synthase (EC 2.7.9.2) (Pyruvate,water dikinase)
DE   (PEP synthase).
GN   PPSA OR DR1727.
OS   Deinococcus radiodurans.
OC   Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
RN   [2]
RP   SEQUENCE OF 259-780 FROM N.A.
RA   Narumi I., Islam S., Cherdchu K., Kikuchi M., Watanabe H.,
RA   Kitayama S., Yamamoto K.;
RT   "IS8301: the second insertion sequence element from Deinococcus
RT   radiodurans.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ATP + pyruvate + H(2)O = AMP +
CC       phosphoenolpyruvate + phosphate.
CC   -!- PATHWAY: ESSENTIAL STEP IN GLUCONEOGENESIS WHEN PYRUVATE AND
CC       LACTATE ARE USED AS A CARBON SOURCE.
CC   -!- SIMILARITY: Belongs to the PEP-utilizing enzyme family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE002014; AAF11283.1; -.
DR   EMBL; AB016803; BAA32387.1; -.
DR   PIR; D75361; D75361.
DR   PIR; T44369; T44369.
DR   TIGR; DR1727; -.
DR   InterPro; IPR008279; PEP_mobile.
DR   InterPro; IPR006318; PEP_P_trans.
DR   InterPro; IPR006319; PEP_synth.
DR   InterPro; IPR000121; PEP_utilizers.
DR   InterPro; IPR002192; PPDK_N_term.
DR   Pfam; PF00391; PEP-utilizers; 1.
DR   Pfam; PF02896; PEP-utilizers_C; 1.
DR   Pfam; PF01326; PPDK_N_term; 1.
DR   PRINTS; PR01736; PHPHTRNFRASE.
DR   ProDom; PD000940; PEP_utilizers; 1.
DR   TIGRFAMs; TIGR01418; PEP_synth; 1.
DR   PROSITE; PS00370; PEP_ENZYMES_PHOS_SITE; 1.
DR   PROSITE; PS00742; PEP_ENZYMES_2; 1.
KW   Transferase; Kinase; ATP-binding; Phosphorylation; Complete proteome.
FT   ACT_SITE    409    409       TELE-PHOSPHOHISTIDINE INTERMEDIATE
FT                                (BY SIMILARITY).
FT   MOD_RES     409    409       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   780 AA;  84895 MW;  AD555076324ADA47 CRC64;

  Query Match             36.9%;  Score 48;  DB 1;  Length 780;
  Best Local Similarity   50.0%;  Pred. No. 35;
  Matches    9;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          3 GLPAVVGLSPGEQEYHRG 20
              |:|||||     :| | |
Db        418 GIPAVVGTGNATRELHNG 435


RESULT 5
PURU_ECOL6
```

```
ID   PURU_ECOL6     STANDARD;      PRT;   280 AA.
AC   P38480;
DT   01-OCT-1994 (Rel. 30, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-FH(4)
DE   hydrolase).
GN   PURU OR C1696 OR SF1232 OR S1318.
OS   Escherichia coli O6, and
OS   Shigella flexneri.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=217992, 623;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX   MEDLINE=22388234; PubMed=12471157;
RA   Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA   Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA   Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA   Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT   "Extensive mosaic structure revealed by the complete genome sequence
RT   of uropathogenic Escherichia coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX   MEDLINE=93023838; PubMed=1406252;
RA   Hromockyj A.E., Tucker S.C., Maurelli A.T.;
RT   "Temperature regulation of Shigella virulence: identification of the
RT   repressor gene virR, an analogue of hns, and partial complementation
RT   by tyrosyl transfer RNA (tRNA1(Tyr)).";
RL   Mol. Microbiol. 6:2113-2124(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX   MEDLINE=22272406; PubMed=12384590;
RA   Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA   Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA   Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA   Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA   Yu J.;
RT   "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT   through comparison with genomes of Escherichia coli K12 and O157.";
RL   Nucleic Acids Res. 30:4432-4441(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX   MEDLINE=22590274; PubMed=12704152;
RA   Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA   Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA   Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA   Schwartz D.C., Blattner F.R.;
RT   "Complete genome sequence and comparative genomics of Shigella
RT   flexneri serotype 2a strain 2457T.";
RL   Infect. Immun. 71:2775-2786(2003).
RN   [5]
RP   IDENTIFICATION.
RX   MEDLINE=94042872; PubMed=8226647;
RA   Nagy P.L., McCorkle G., Zalkin H.;
RT   "purU, a source of formate for purT-dependent phosphoribosyl-N-
RT   formylglycinamide synthesis.";
RL   J. Bacteriol. 175:7066-7073(1993).
CC   -!- FUNCTION: PRODUCES FORMATE FROM FORMYL-TETRAHYDROFOLATE. PROVIDES
CC       THE MAJOR SOURCE OF FORMATE FOR THE PURT-DEPENDENT SYNTHESIS OF
CC       5'-PHOSPHORIBOSYL-N-FORMYLGLYCINAMIDE (FGAR) DURING AEROBIC
CC       GROWTH. HAS A ROLE IN REGULATING THE ONE-CARBON POOL.
CC   -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + H(2)O = formate +
CC       tetrahydrofolate.
CC   -!- ENZYME REGULATION: Activated by methionine, inhibited by glycine
CC       (By similarity).
CC   -!- PATHWAY: De novo purine biosynthesis.
CC   -!- SUBUNIT: Homohexamer (By similarity).
CC   -!- SIMILARITY: SOME, TO GAR TRANSFORMYLASE (PURN).
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE016760; AAN80163.1; -.
DR   EMBL; X66849; -; NOT_ANNOTATED_CDS.
DR   EMBL; AE015150; AAN42845.1; -.
DR   EMBL; AE016982; AAP16730.1; -.
DR   HSSP; P08179; 1GRC.
DR   InterPro; IPR002912; ACT.
DR   InterPro; IPR002376; formyl_transf.
DR   InterPro; IPR004810; PurU.
DR   Pfam; PF01842; ACT; 1.
DR   Pfam; PF00551; formyl_transf; 1.
DR   PRINTS; PR01575; FFH4HYDRLASE.
DR   TIGRFAMs; TIGR00655; PurU; 1.
KW   Purine biosynthesis; Hydrolase; One-carbon metabolism;
KW   Complete proteome.
FT   ACT_SITE    225    225       BY SIMILARITY.
FT   CONFLICT     44     44       R -> L (IN REF. 2).
SQ   SEQUENCE   280 AA;  31920 MW;  55BC16B62727A419 CRC64;

  Query Match             36.2%;  Score 47;  DB 1;  Length 280;
  Best Local Similarity   52.9%;  Pred. No. 17;
  Matches    9;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :|  |  | | ||
Db        193 LPAFIGARPYHQAYERG 209


RESULT 6
PURU_ECOLI
ID   PURU_ECOLI     STANDARD;      PRT;   280 AA.
AC   P37051;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-FH(4)
DE   hydrolase).
GN   PURU OR TGS OR B1232.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-6.
RC   STRAIN=K12;
RX   MEDLINE=94042872; PubMed=8226647;
RA   Nagy P.L., McCorkle G., Zalkin H.;
RT   "purU, a source of formate for purT-dependent phosphoribosyl-N-
RT   formylglycinamide synthesis.";
RL   J. Bacteriol. 175:7066-7073(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=94110230; PubMed=8282700;
RA   Boesl M., Kersten H.;
RT   "Organization and functions of genes in the upstream region of tyrT
RT   of Escherichia coli: phenotypes of mutants with partial deletion of a
RT   new gene (tgs).";
RL   J. Bacteriol. 176:221-231(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
```

```
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [5]
RP   CHARACTERIZATION.
RX   MEDLINE=95173107; PubMed=7868604;
RA   Nagy P.L., Marolewski A., Benkovic S.J., Zalkin H.;
RT   "Formyltetrahydrofolate hydrolase, a regulatory enzyme that functions
RT   to balance pools of tetrahydrofolate and one-carbon tetrahydrofolate
RT   adducts in Escherichia coli.";
RL   J. Bacteriol. 177:1292-1298(1995).
CC   -!- FUNCTION: PRODUCES FORMATE FROM FORMYL-TETRAHYDROFOLATE. PROVIDES
CC       THE MAJOR SOURCE OF FORMATE FOR THE PURT-DEPENDENT SYNTHESIS OF
CC       5'-PHOSPHORIBOSYL-N-FORMYLGLYCINAMIDE (FGAR) DURING AEROBIC
CC       GROWTH. HAS A ROLE IN REGULATING THE ONE-CARBON POOL.
CC   -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + H(2)O = formate +
CC       tetrahydrofolate.
CC   -!- ENZYME REGULATION: Activated by methionine, inhibited by glycine.
CC   -!- PATHWAY: De novo purine biosynthesis.
CC   -!- SUBUNIT: Homohexamer.
CC   -!- SIMILARITY: SOME, TO GAR TRANSFORMYLASE (PURN).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L20251; AAC36846.1; -.
DR   EMBL; M64675; AAA16860.1; ALT_INIT.
DR   EMBL; AE000221; AAC74314.1; -.
DR   EMBL; D90758; BAA36100.1; -.
DR   EMBL; D90759; BAA36112.1; -.
DR   EMBL; D90852; BAA16026.1; -.
DR   PIR; C36871; C36871.
DR   EcoGene; EG11819; purU.
DR   InterPro; IPR002912; ACT.
DR   InterPro; IPR002376; formyl_transf.
DR   InterPro; IPR004810; PurU.
DR   Pfam; PF01842; ACT; 1.
DR   Pfam; PF00551; formyl_transf; 1.
DR   PRINTS; PR01575; FFH4HYDRLASE.
DR   TIGRFAMs; TIGR00655; PurU; 1.
KW   Purine biosynthesis; Hydrolase; One-carbon metabolism;
KW   Complete proteome.
FT   ACT_SITE    225    225       BY SIMILARITY.
SQ   SEQUENCE   280 AA;  31934 MW;  5667406D2727A2C2 CRC64;

  Query Match             36.2%;  Score 47;  DB 1;  Length 280;
  Best Local Similarity   52.9%;  Pred. No. 17;
  Matches    9;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :|  |  | | ||
Db        193 LPAFIGARPYHQAYERG 209
```

```
RESULT 7
ARAB_VIBPA
ID   ARAB_VIBPA     STANDARD;      PRT;   567 AA.
AC   Q87FK5;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   L-ribulokinase (EC 2.7.1.16).
GN   ARAB OR VPA1674.
OS   Vibrio parahaemolyticus.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=670;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIMD 2210633 / Serotype O3:K6;
RX   MEDLINE=22508454; PubMed=12620739;
RA   Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA   Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA   Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT   "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT   distinct from that of V. cholerae.";
RL   Lancet 361:743-749(2003).
CC   -!- CATALYTIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-
CC       phosphate.
CC   -!- PATHWAY: L-arabinose catabolism; second step.
CC   -!- SIMILARITY: Belongs to the ribulokinase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP005089; BAC63017.1; -.
DR   HAMAP; MF_00520; -; 1.
DR   InterPro; IPR000577; FGGY_kin.
DR   Pfam; PF00370; FGGY; 1.
DR   Pfam; PF02782; FGGY_C; 1.
KW   Transferase; Kinase; Arabinose catabolism; Complete proteome.
SQ   SEQUENCE   567 AA;  61788 MW;  B303B547EBEA9B8B CRC64;

  Query Match             36.2%;  Score 47;  DB 1;  Length 567;
  Best Local Similarity   42.9%;  Pred. No. 35;
  Matches    9;  Conservative    4;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          3 GLPAVVGLSPGEQEYHRGGVG 23
              |||  : :: || : | | ||
Db        267 GLPEGIAIAIGEFDCHMGAVG 287


RESULT 8
CADM_MOUSE
ID   CADM_MOUSE     STANDARD;      PRT;   813 AA.
AC   Q9WTP5;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cadherin-22 precursor (PB-cadherin).
GN   CDH22.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGE.
RC   STRAIN=ICR; TISSUE=Brain;
RX   MEDLINE=99326347; PubMed=10398531;
RA   Kitajima K., Koshimizu U., Nakamura T.;
RT   "Expression of a novel type of classic cadherin, PB-cadherin in
```

```
RT   developing brain and limb buds.";
RL   Dev. Dyn. 215:206-214(1999).
CC   -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC       They preferentially interact with themselves in a homophilic
CC       manner in connecting cells; cadherins may thus contribute to the
CC       sorting of heterogeneous cell types. PB-cadherins may have a role
CC       in the morphological organization of pituitary gland and brain
CC       tissues.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: Predominantly expressed in brain. Abundant in
CC       olfactory bulb, cerebrum, and cerebellum, less in pons, medulla,
CC       and spinal cord. Low expression in heart. No expression in lung,
CC       liver, spleen, kidney, testis, stomach, intestine, colon, and
CC       placenta.
CC   -!- DEVELOPMENTAL STAGE: Expressed at 9.5 dpc onwards. At 10.5 dpc, in
CC       brain (telencephalic vesicles and isthmus), spinal cord and limb
CC       buds (in the zone of polarizing activity). At 14.5 dpc, in
CC       olfactory bulb and cerebellum.
CC   -!- INDUCTION: Down-regulated by thyroid hormone.
CC   -!- SIMILARITY: Contains 5 cadherin domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB019618; BAA34426.1; -.
DR   MGD; MGI:1341843; Cdh22.
DR   HSSP; P15116; 1NCJ.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000233; Cadherin_C_term.
DR   Pfam; PF00028; cadherin; 5.
DR   Pfam; PF01049; Cadherin_C_term; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 5.
DR   PROSITE; PS00232; CADHERIN_1; 2.
DR   PROSITE; PS50268; CADHERIN_2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW   Signal.
FT   SIGNAL        1     33       POTENTIAL.
FT   CHAIN        34    813       CADHERIN-22.
FT   DOMAIN       33    621       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    622    642       POTENTIAL.
FT   DOMAIN      643    813       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       61    165       CADHERIN 1.
FT   DOMAIN      166    274       CADHERIN 2.
FT   DOMAIN      275    391       CADHERIN 3.
FT   DOMAIN      392    495       CADHERIN 4.
FT   DOMAIN      496    613       CADHERIN 5.
FT   CARBOHYD    159    159       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    463    463       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    609    609       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   813 AA;  88021 MW;  5510F9848D976567 CRC64;

  Query Match             36.2%;  Score 47;  DB 1;  Length 813;
  Best Local Similarity   50.0%;  Pred. No. 51;
  Matches   11;  Conservative    2;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          2 AGLPAVVGLSPGEQEYHRGGVG 23
              |  ||   |||| || :: | |
Db         34 ASTPAPSSLSPGAQEDNQLGAG 55


RESULT 9
PGCA_CHICK
ID   PGCA_CHICK     STANDARD;      PRT;  2109 AA.
AC   P07898; Q90810; Q90820; Q90991; Q91047;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Aggrecan core protein precursor (Cartilage-specific proteoglycan core
DE   protein) (CSPCP).
GN   AGC1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=White leghorn; TISSUE=Embryo;
RX   MEDLINE=94043149; PubMed=8226878;
RA   Li H., Schwartz N.B., Vertel B.M.;
RT   "cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core
RT   protein and identification of a stop codon in the aggrecan gene
RT   associated with the chondrodystrophy, nanomelia.";
RL   J. Biol. Chem. 268:23504-23511(1993).
RN   [2]
RP   SEQUENCE OF 1042-1559 FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=90307744; PubMed=1694853;
RA   Krueger R.C. Jr., Fields T.A., Mensch J.R. Jr., Schwartz N.B.;
RT   "Chick cartilage chondroitin sulfate proteoglycan core protein. II.
RT   Nucleotide sequence of cDNA clone and localization of the S103L
RT   epitope.";
RL   J. Biol. Chem. 265:12088-12097(1990).
RN   [3]
RP   SEQUENCE OF 1-1855 AND 1893-2109 FROM N.A.
RC   TISSUE=Cartilage;
RX   MEDLINE=93111968; PubMed=1339285;
RA   Chandrasekaran L., Tanzer M.L.;
RT   "Molecular cloning of chicken aggrecan. Structural analyses.";
RL   Biochem. J. 288:903-910(1992).
RN   [4]
RP   ERRATUM.
RX   MEDLINE=94107258; PubMed=8280087;
RA   Chandrasekaran L., Tanzer M.L.;
RL   Biochem. J. 296:885-887(1993).
RN   [5]
RP   SEQUENCE OF 1492-1610 FROM N.A.
RC   STRAIN=White leghorn; TISSUE=Chondrocytes;
RX   MEDLINE=95128519; PubMed=7827752;
RA   Primorac D., Stover M.L., Clark S.H., Rowe D.W.;
RT   "Molecular basis of nanomelia, a heritable chondrodystrophy of
RT   chicken.";
RL   Matrix Biol. 14:297-305(1994).
RN   [6]
RP   SEQUENCE OF 1894-2109 FROM N.A.
RX   MEDLINE=89008500; PubMed=3170613;
RA   Tanaka T., Har-El R., Tanzer M.L.;
RT   "Partial structure of the gene for chicken cartilage proteoglycan
RT   core protein.";
RL   J. Biol. Chem. 263:15831-15835(1988).
RN   [7]
RP   SEQUENCE OF 1693-1855 AND 1893-2109 FROM N.A.
RX   MEDLINE=86259736; PubMed=3460082;
RA   Sai S., Tanaka T., Kosher R.A., Tanzer M.L.;
RT   "Cloning and sequence analysis of a partial cDNA for chicken
RT   cartilage proteoglycan core protein.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:5081-5085(1986).
CC   -!- FUNCTION: This proteoglycan is a major component of extracellular
CC       matrix of cartilagenous tissues. A major function of this protein
CC       is to resist compression in cartilage. It binds avidly to
CC       hyaluronic acid via an amino-terminal globular region. May play a
CC       regulatory role in the matrix assembly of the cartilage.
CC   -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
CC       similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=P07898-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P07898-2; Sequence=VSP_003073;
```

```
CC   -!- DOMAIN: Two globular domains, G1 and G2, comprise the amino
CC       terminus of the proteoglycan, while another globular region, G3,
CC       makes up the COOH terminus. G1 contains link domains and thus
CC       consists of three disulfide-bonded loop structures designated as
CC       the A, B, B' motifs. G2 is similar to G1. The keratan sulfate (KS)
CC       and the chondroitin sulfate (CS) attachment domains lie between G2
CC       and G3.
CC   -!- PTM: Contains mostly chondroitin sulfate, but also keratan sulfate
CC       chains, N-linked and O-linked oligosaccharides.
CC   -!- DISEASE: DEFECTS IN AGC1 ARE THE CAUSE OF NANOMELIA, A LETHAL
CC       CONNECTIVE TISSUE DISORDER AFFECTING CARTILAGE DEVELOPMENT
CC       (CHONDRODYSTROPHY) CHARACTERIZED BY SHORTENED AND MALFORMED LIMBS.
CC       AGGRECAN IS TRUNCATED AT ITS C-TERMINAL IN THE CS-2 BINDING DOMAIN
CC       AND IS NOT ANYMORE SECRETED FROM THE CHONDROCYTES.
CC   -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC   -!- SIMILARITY: Contains 4 link domains.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L21913; AAB19128.1; -.
DR   EMBL; M38187; AAA48731.1; -.
DR   EMBL; M88101; -; NOT_ANNOTATED_CDS.
DR   EMBL; S74657; AAC60751.1; -.
DR   EMBL; S74656; AAC60751.1; JOINED.
DR   EMBL; J04028; AAA48719.1; -.
DR   EMBL; M13993; AAA48720.1; -.
DR   PIR; I50421; I50421.
DR   HSSP; P08709; 1BF9.
DR   InterPro; IPR002353; AntifreezeII.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR001304; Lectin_C.
DR   InterPro; IPR000538; Link.
DR   InterPro; IPR003324; SGXXSG.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF02339; SGXXSG; 56.
DR   Pfam; PF00084; sushi; 1.
DR   Pfam; PF00193; Xlink; 4.
DR   PRINTS; PR00356; ANTIFREEZEII.
DR   PRINTS; PR01265; LINKMODULE.
DR   ProDom; PD000918; Link; 4.
DR   SMART; SM00032; CCP; 1.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00445; LINK; 4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS50026; EGF_3; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
DR   PROSITE; PS01241; LINK; 4.
KW   Glycoprotein; Proteoglycan; Lectin; Signal; Sushi; EGF-like domain;
KW   Alternative splicing; Repeat; Immunoglobulin domain.
```

```
FT   SIGNAL       1    16       POTENTIAL.
FT   CHAIN       17  2109       AGGRECAN CORE PROTEIN.
FT   DOMAIN      34   143       IG-LIKE V-TYPE.
FT   DOMAIN     166   243       LINK 1.
FT   DOMAIN     264   346       LINK 2.
FT   DOMAIN     537   614       LINK 3.
FT   DOMAIN     635   716       LINK 4.
FT   DOMAIN    1363  1742       19 X 20 AA TANDEM-REPEAT.
FT   DOMAIN    1855  1892       EGF-LIKE.
FT   DOMAIN    1901  2019       C-TYPE LECTIN.
FT   DOMAIN    2023  2081       SUSHI.
FT   DOMAIN      48   137       G1-A.
FT   DOMAIN     148   243       G1-B.
FT   DOMAIN     249   346       G1-B'.
FT   DOMAIN     519   613       G2-B.
FT   DOMAIN     620   715       G2-B'.
FT   DOMAIN     718   803       KS.
FT   DOMAIN     805  1264       CS-1.
FT   DOMAIN    1265  1742       CS-2.
FT   DOMAIN    1893  2109       G3.
FT   DISULFID    51   129       BY SIMILARITY.
FT   DISULFID   171   242       BY SIMILARITY.
FT   DISULFID   195   216       BY SIMILARITY.
FT   DISULFID   269   345       BY SIMILARITY.
FT   DISULFID   293   314       BY SIMILARITY.
FT   DISULFID   542   613       BY SIMILARITY.
FT   DISULFID   566   587       BY SIMILARITY.
FT   DISULFID   640   715       BY SIMILARITY.
FT   DISULFID   664   685       BY SIMILARITY.
FT   DISULFID  1859  1870       BY SIMILARITY.
FT   DISULFID  1864  1879       BY SIMILARITY.
FT   DISULFID  1881  1890       BY SIMILARITY.
FT   DISULFID  1897  1908       BY SIMILARITY.
FT   DISULFID  1925  2017       BY SIMILARITY.
FT   DISULFID  1993  2009       BY SIMILARITY.
FT   DISULFID  2024  2067       BY SIMILARITY.
FT   DISULFID  2053  2080       BY SIMILARITY.
FT   CARBOHYD    76    76       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   122   122       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   330   330       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   388   388       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   439   439       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   644   644       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   700   700       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   765   765       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   801   801       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC  1856  1892       Missing (in isoform 2).
FT                              /FTId=VSP_003073.
FT   CONFLICT   362   362       E -> D (IN REF. 3).
```

```
  Query Match           36.2%;  Score 47;  DB 1;  Length 2109;
  Best Local Similarity 47.6%;  Pred. No. 1.4e+02;
  Matches   10;  Conservative    3;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHRGGVGV 24
              |  ||  :|| ||  :| :||
Db        957 LVEVVTAAPGRQEEGKGSIGV 977
```

```
RESULT 10
COBI_MYCTU
ID   COBI_MYCTU     STANDARD;      PRT;   508 AA.
AC   Q10677;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Cobalamin biosynthesis protein COBIJ [Includes: Precorrin-2 C20-
DE   methyltransferase (EC 2.1.1.130) (S-adenosyl-L-methionine--precorrin-2
DE   methyltransferase) (SP2MT); Precorrin-3 methylase (EC 2.1.1.-)].
GN   COBIJ OR COBI OR RV2066 OR MT2126 OR MTCY49.05 OR MB2092.
OS   Mycobacterium tuberculosis, and
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
```

```
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773, 1765;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX   MEDLINE=22206494; PubMed=12218036;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT   "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   J. Bacteriol. 184:5479-5490(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.bovis; STRAIN=AF2122/97;
RX   MEDLINE=22709107; PubMed=12788972;
RA   Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA   Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA   Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA   Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT   "The complete genome sequence of Mycobacterium bovis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC   -!- FUNCTION: METHYLATES PRECORRIN-2 AT THE C-20 POSITION TO PRODUCE
CC       PRECORRIN-3A (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-2 = S-
CC       adenosyl-L-homocysteine + precorrin-3A.
CC   -!- PATHWAY: Cobalamin biosynthesis.
CC   -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS SUMT, CYSG, CBIF/COBM
CC       AND CBIL/COBI.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z73966; CAA98214.1; -.
DR   EMBL; AE007063; AAK46406.1; -.
DR   EMBL; BX248341; CAD96945.1; -.
DR   PIR; E70764; E70764.
DR   TIGR; MT2126; -.
DR   TubercuList; Rv2066; -.
DR   InterPro; IPR006364; CobI_CbiL.
DR   InterPro; IPR006363; CobJ.
DR   InterPro; IPR000878; Cor/por_Metransf.
DR   InterPro; IPR003043; Uropor_Metransf.
DR   Pfam; PF00590; TP_methylase; 2.
DR   TIGRFAMs; TIGR01467; cobI_cbiL; 1.
DR   TIGRFAMs; TIGR01466; cobJ_cbiH; 1.
DR   PROSITE; PS00839; SUMT_1; 1.
DR   PROSITE; PS00840; SUMT_2; 1.
KW   Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW   Methyltransferase; Multifunctional enzyme; Complete proteome.
FT   DOMAIN       1    243       PRECORRIN-2 C20-METHYLTRANSFERASE.
FT   DOMAIN     244    508       PRECORRIN-3 METHYLASE.
SQ   SEQUENCE   508 AA;  53910 MW;  95AC066F022C4DC1 CRC64;

  Query Match             35.4%;  Score 46;  DB 1;  Length 508;
  Best Local Similarity   47.1%;  Pred. No. 44;
  Matches    8;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 MAGLPAVVGLSPGEQEY 17
              : |  ||||| ||: ::
Db        245 LTGTVAVVGLGPGDSDW 261


RESULT 11
CNA2_MOUSE
ID   CNA2_MOUSE     STANDARD;      PRT;  1172 AA.
AC   Q80XC6; Q8R3D7; Q99LT9;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Protein C14orf102 homolog.
GN   C14ORF102.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   CONCEPTUAL TRANSLATION, AND RECONSTRUCTION FROM EST.
RA   Reynaud S.;
RL   Unpublished observations (AUG-2003).
RN   [2]
RP   SEQUENCE OF 21-1172 FROM N.A.
RC   TISSUE=Breast tumor, and Olfactory epithelium;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SIMILARITY: Contains 5 HAT repeats.
CC   -!- CAUTION: This is a conceptual translation; a frameshift was
CC       introduced in position 28 to produce the correct N-terminus, and
CC       to extend the similarity with the human ortholog.
CC   -!- CAUTION: Ref.1 (AAH51175) sequence differs from that shown due to
CC       a stop codon in position 1146 which was translated as Gln to
CC       extend the sequence.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; BY182441; -; NOT_ANNOTATED_CDS.
DR   EMBL; BC002230; AAH02230.1; -.
DR   EMBL; BC025577; AAH25577.1; ALT_INIT.
DR   EMBL; BC051175; AAH51175.1; ALT_SEQ.
```

```
DR   InterPro; IPR003107; HAT.
DR   SMART; SM00386; HAT; 5.
KW   Repeat; Coiled coil.
FT   DOMAIN       78    330       COILED COIL (POTENTIAL).
FT   REPEAT      314    346       HAT 1.
FT   REPEAT      404    436       HAT 2.
FT   REPEAT      766    800       HAT 3.
FT   REPEAT      986   1018       HAT 4.
FT   REPEAT     1075   1109       HAT 5.
FT   DOMAIN       86     93       POLY-LYS.
FT   DOMAIN      146    154       POLY-ALA.
FT   CONFLICT    707    707       R -> H (IN REF. 1; AAH25577).
FT   CONFLICT    741    742       LS -> VC (IN REF. 1; AAH25577).
SQ   SEQUENCE   1172 AA;  133424 MW;  ABE37A8BCE7A528F CRC64;

  Query Match             35.4%;  Score 46;  DB 1;  Length 1172;
  Best Local Similarity   42.9%;  Pred. No. 1e+02;
  Matches    9;  Conservative    3;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          5 PAVVGLSPGEQEYHRGGVGVL 25
              | :  |  |||| ||  : :|
Db        351 PGIYALGEGEQEKHRKSLKLL 371


RESULT 12
PTSB_VIBAL
ID   PTSB_VIBAL     STANDARD;      PRT;   479 AA.
AC   P22825;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   PTS system, sucrose-specific IIBC component (EIIBC-SCR) (Sucrose-
DE   permease IIBC component) (Phosphotransferase enzyme II, BC component)
DE   (EC 2.7.1.69) (EII-SCR).
GN   SCRA.
OS   Vibrio alginolyticus.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=663;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91071601; PubMed=2174811;
RA   Blatch G.L., Scholle R.R., Woods D.R.;
RT   "Nucleotide sequence and analysis of the Vibrio alginolyticus sucrose
RT   uptake-encoding region.";
RL   Gene 95:17-23(1990).
RN   [2]
RP   SEQUENCE OF 1-12 FROM N.A.
RX   MEDLINE=91285433; PubMed=2060795;
RA   Blatch G.L., Woods D.R.;
RT   "Nucleotide sequence and analysis of the Vibrio alginolyticus scr
RT   repressor-encoding gene (scrR).";
RL   Gene 101:45-50(1991).
CC   -!- FUNCTION: This is a component of the phosphoenolpyruvate-dependent
CC       sugar phosphotransferase system (PTS), a major carbohydrate active
CC       -transport system. The IICD domains contain the sugar binding site
CC       and the transmembrane channel; the IIA domain contains the primary
CC       phosphorylation site (the donor is phospho-HPr); IIA transfers its
CC       phosphoryl group to the IIB domain which finally transfers it to
CC       the sugar.
CC   -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC       histidine + sugar phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- SIMILARITY: Contains 1 PTS EIIB domain.
CC   -!- SIMILARITY: Contains 1 PTS EIIC domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M76768; AAA27555.1; -.
DR   EMBL; M35009; AAA27557.2; -.
DR   PIR; JQ0781; JQ0781.
DR   HSSP; P05053; 1IBA.
DR   InterPro; IPR001996; Ptrans_EIIB.
DR   InterPro; IPR003352; Ptrans_EIIC.
DR   InterPro; IPR004719; PTSIIC_glc.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD001476; PTS_EIIB; 1.
DR   TIGRFAMs; TIGR00826; EIIB_glc; 1.
DR   TIGRFAMs; TIGR00852; pts-Glc; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Phosphotransferase system; Sugar transport; Transferase;
KW   Transmembrane; Inner membrane; Phosphorylation.
FT   DOMAIN       1      ?       EIIB.
FT   DOMAIN       ?    479       EIIC.
FT   MOD_RES     26     26       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES    324    324       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   479 AA;  49890 MW;  FBF906B5170E3EB7 CRC64;

  Query Match             34.6%;  Score 45;  DB 1;  Length 479;
  Best Local Similarity   33.3%;  Pred. No. 58;
  Matches    7;  Conservative    7;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          2 AGLPAVVGLSPGEQEYHRGGV 22
              |||| :: ::| :  |:  |:
Db        431 AGLPGIISITPDKIGYYIAGM 451


RESULT 13
ACON_MYCAV
ID   ACON_MYCAV     STANDARD;      PRT;   961 AA.
AC   O08451;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Aconitate hydratase (EC 4.2.1.3) (Citrate hydro-lyase) (Aconitase).
GN   ACN.
OS   Mycobacterium avium.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1764;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GIR10;
RX   MEDLINE=98195739; PubMed=9534249;
RA   Labo M., Gusberti L., de Rossi E., Speziale P., Riccardi G.;
RT   "Determination of a 15437 bp nucleotide sequence around the inhA gene
RT   of Mycobacterium avium and similarity analysis of the products of
RT   putative ORFs.";
RL   Microbiology 144:807-814(1998).
CC   -!- CATALYTIC ACTIVITY: Citrate = cis-aconitate + H(2)O.
CC   -!- COFACTOR: Binds 1 4Fe-4S cluster per subunit (By similarity).
CC   -!- PATHWAY: Tricarboxylic acid cycle.
CC   -!- SUBUNIT: Monomer (By similarity).
CC   -!- SIMILARITY: Belongs to the aconitase/IPM isomerase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF002133; AAC46192.1; -.
DR   InterPro; IPR006249; Aconitase_1.
DR   InterPro; IPR000573; Aconitase_C.
DR   InterPro; IPR001030; Aconitase_N.
DR   Pfam; PF00330; aconitase; 1.
DR   Pfam; PF00694; Aconitase_C; 1.
```

```
DR   PRINTS; PR00415; ACONITASE.
DR   ProDom; PD000511; Aconitase_N; 1.
DR   TIGRFAMs; TIGR01341; aconitase_1; 1.
DR   PROSITE; PS00450; ACONITASE_1; 1.
DR   PROSITE; PS01244; ACONITASE_2; 1.
KW   Lyase; Tricarboxylic acid cycle; Iron-sulfur; 4Fe-4S.
FT   METAL       499    499       IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT   METAL       565    565       IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT   METAL       568    568       IRON-SULFUR (4FE-4S) (BY SIMILARITY).
SQ   SEQUENCE   961 AA;  104025 MW;  689E66D85E31F596 CRC64;

  Query Match             34.6%;  Score 45;  DB 1;  Length 961;
  Best Local Similarity   48.0%;  Pred. No. 1.2e+02;
  Matches   12;  Conservative    5;  Mismatches    6;  Indels    2;  Gaps    2;

Qy          2 AGLPAVVGL-SPGEQEYHRGGVGVL 25
              |   ||| : :||| :|:| | |:|
Db        927 AEFDAVVRIDTPGEADYYRNG-GIL 950


RESULT 14
ILVD_CHRVO
ID   ILVD_CHRVO     STANDARD;      PRT;   618 AA.
AC   Q7NYJ7;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Dihydroxy-acid dehydratase (EC 4.2.1.9) (DAD).
GN   ILVD OR CV1277.
OS   Chromobacterium violaceum.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC   Neisseriaceae; Chromobacterium.
OX   NCBI_TaxID=536;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 12472 / DSM 30191;
RX   MEDLINE=22882880; PubMed=14500782;
RA   Vasconcelos A.T.R., de Almeida D.F., Hungria M., Guimaraes C.T.,
RA   Antonio R.V., Almeida F.C., de Almeida L.G.P., de Almeida R.,
RA   Alves-Gomes J.A., Andrade E.M., Araripe J., de Araujo M.F.F.,
RA   Astolfi-Filho S., Azevedo V., Baptista A.J., Bataus L.A.M.,
RA   Batista J.S., Belo A., van den Berg C., Bogo M., Bonatto S.,
RA   Bordignon J., Brigido M.M., Brito C.A., Brocchi M., Burity H.A.,
RA   Camargo A.A., Cardoso D.D.P., Carneiro N.P., Carraro D.M.,
RA   Carvalho C.M.B., Cascardo J.C.M., Cavada B.S., Chueire L.M.O.,
RA   Creczynski-Pasa T.B., Cunha-Junior N.C., Fagundes N., Falcao C.L.,
RA   Fantinatti F., Farias I.P., Felipe M.S.S., Ferrari L.P., Ferro J.A.,
RA   Ferro M.I.T., Franco G.R., Freitas N.S.A., Furlan L.R.,
RA   Gazzinelli R.T., Gomes E.A., Goncalves P.R., Grangeiro T.B.,
RA   Grattapaglia D., Grisard E.C., Hanna E.S., Jardim S.N., Laurino J.,
RA   Leoi L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.C.P.,
RA   Madeira H.M.F., Manfio G.P., Maranhao A.Q., Martins W.S.,
RA   di Mauro S.M.Z., de Medeiros S.R.B., Meissner R.V., Moreira M.A.M.,
RA   Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,
RA   Paixao R.F.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,
RA   Pereira M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Potrich D.P.,
RA   Ramalho-Neto C.E., Reis A.M.M., Rigo L.U., Rondinelli E.,
RA   Santos E.B.P., Santos F.R., Schneider M.P.C., Seuanez H.N.,
RA   Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,
RA   Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,
RA   Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Urmenyi T.,
RA   Vettore A., Wassem R., Zaha A., Simpson A.J.G.;
RT   "The complete genome sequence of Chromobacterium violaceum reveals
RT   remarkable and exploitable bacterial adaptability.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).
CC   -!- CATALYTIC ACTIVITY: 2,3-dihydroxy-3-methylbutanoate = 3-methyl-2-
CC       oxobutanoate + H(2)O.
CC   -!- COFACTOR: Binds 1 4Fe-4S cluster (Potential).
CC   -!- PATHWAY: Valine and isoleucine biosynthesis; fourth step.
CC   -!- SIMILARITY: Belongs to the ilvD / edd family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE016914; AAQ58952.1; -.
DR   HAMAP; MF_00012; -; 1.
DR   PROSITE; PS00886; ILVD_EDD_1; 1.
DR   PROSITE; PS00887; ILVD_EDD_2; 1.
KW   Branched-chain amino acid biosynthesis; Lyase; Iron; Iron-sulfur;
KW   4Fe-4S; Complete proteome.
FT   METAL       122    122       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       195    195       IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ   SEQUENCE   618 AA;  65582 MW;  3273399ABA192A05 CRC64;

  Query Match             34.2%;  Score 44.5;  DB 1;  Length 618;
  Best Local Similarity   30.0%;  Pred. No. 89;
  Matches    9;  Conservative    7;  Mismatches    7;  Indels    7;  Gaps    1;

Qy          3 GLPAVVGLSPGEQEYHR-------GGVGVL 25
              |:| :  ::|  |:||         | :|:|
Db        307 GVPCLSKVAPATQKYHMEDVHRAGGVIGIL 336


RESULT 15
M18A_HUMAN
ID   M18A_HUMAN     STANDARD;      PRT;  2054 AA.
AC   Q92614; Q8IXP8;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Myosin XVIIIA (Myosin 18A) (Myosin containing PDZ domain).
GN   MYO18A OR MYSPDZ OR KIAA0216.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain;
RX   MEDLINE=97191544; PubMed=9039502;
RA   Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayasi Y.,
RA   Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. VI.
RT   The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT   analysis of cDNA clones from cell line KG-1 and brain.";
RL   DNA Res. 3:321-329(1996).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RC   TISSUE=Testis;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
```

```
RP   ALTERNATIVE SPLICING (ISOFORMS 1 AND 2).
RX   MEDLINE=22646225; PubMed=12761286;
RA   Mori K., Furusawa T., Okubo T., Inoue T., Ikawa S., Yanai N.,
RA   Mori K.J., Obinata M.;
RT   "Genome structure and differential expression of two isoforms of a
RT   novel PDZ-containing myosin (MysPDZ) (Myo18A).";
RL   J. Biochem. 133:405-413(2003).
CC   -!- FUNCTION: May be involved in the maintenance of the stromal cell
CC       architectures required for cell to cell contact (By similarity).
CC   -!- SUBCELLULAR LOCATION: Isoform 1 co-localizes with the
CC       endoplasmatic reticulum-Golgi complex; isoform 2, which lacks the
CC       PDZ domain, is diffusely localized in the cytoplasm (By
CC       similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=1; Synonyms=Alpha;
CC         IsoId=Q92614-1; Sequence=Displayed;
CC       Name=2; Synonyms=Beta;
CC         IsoId=Q92614-2; Sequence=VSP_007869, VSP_007870;
CC       Name=3;
CC         IsoId=Q92614-3; Sequence=VSP_007871, VSP_007872;
CC         Note=No experimental confirmation available;
CC   -!- SIMILARITY: Contains 1 IQ domain.
CC   -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC   -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC   -!- CAUTION: The TIAF1 protein is coded in the 3'UTR region of MysPDZ.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D86970; BAA13206.2; ALT_INIT.
DR   EMBL; BC039612; AAH39612.1; -.
DR   PIR; PT0271; PT0271.
DR   GO; GO:0005653; C:perinuclear space; TAS.
DR   GO; GO:0006916; P:anti-apoptosis; TAS.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR003345; M_repeat.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR001478; PDZ.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF02370; M; 7.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00595; PDZ; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS50096; IQ; 1.
DR   PROSITE; PS50106; PDZ; 1.
KW   Myosin; ATP-binding; Coiled coil; Alternative splicing.
FT   DOMAIN      220    311       PDZ.
FT   DOMAIN      420   1186       MYOSIN HEAD-LIKE.
FT   DOMAIN     1188   1217       IQ.
FT   DOMAIN     1246   1971       COILED COIL (POTENTIAL).
FT   NP_BIND     498    505       ATP (POTENTIAL).
FT   VARSPLIC      1    331       Missing (in isoform 2).
FT                                /FTId=VSP_007869.
FT   VARSPLIC    332    333       SD -> MR (in isoform 2).
FT                                /FTId=VSP_007870.
FT   VARSPLIC   1571   1607       Missing (in isoform 3).
FT                                /FTId=VSP_007871.
FT   VARSPLIC   1952   1966       Missing (in isoform 3).
FT                                /FTId=VSP_007872.
SQ   SEQUENCE   2054 AA;  233113 MW;  52BFA0AA273E18F7 CRC64;

  Query Match             34.2%;  Score 44.5;  DB 1;  Length 2054;
  Best Local Similarity   38.5%;  Pred. No. 3.1e+02;

  Matches   10;  Conservative    3;  Mismatches    4;  Indels    9;  Gaps    1;

Qy          7 VVGLSPGEQE---------YHRGGVG 23
              |:|:|| ||:          || |  |
Db        651 VLGISPDEQKACWFILAAIYHLGAAG 676


Search completed: May  7, 2004, 12:34:30
Job time : 11 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:26:40 ; Search time 46.75 Seconds
(without alignments)
168.726 Million cell updates/sec

Title: US-09-786-214A-5
Perfect score: 130
Sequence: 1 MAGLPAVVGLSPGEQEYHRGGVGVL 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	67.5	51.9	41	11	Q8K408	Q8k408 rattus norv
2	56	43.1	242	16	Q89Y25	Q89y25 bradyrhizob
3	54	41.5	330	16	Q87YM5	Q87ym5 pseudomonas
4	52	40.0	150	16	Q8PE55	Q8pe55 xanthomonas
5	51	39.2	426	17	Q9YBR0	Q9ybr0 aeropyrum p
6	50.5	38.8	383	11	Q8K2L1	Q8k2l1 mus musculu
7	50.5	38.8	601	16	Q8FVA5	Q8fva5 brucella su
8	50.5	38.8	615	16	Q8YD21	Q8yd21 brucella me
9	50.5	38.8	662	11	Q8VHY3	Q8vhy3 mus musculu
10	50.5	38.8	680	11	Q8BSX1	Q8bsx1 mus musculu
11	50.5	38.8	758	10	Q8L525	Q8l525 oryza sativ
12	50	38.5	277	16	Q9KQK6	Q9kqk6 vibrio chol
13	50	38.5	284	16	Q7VFI5	Q7vfi5 helicobacte
14	50	38.5	335	16	Q82RR7	Q82rr7 streptomyce
15	50	38.5	381	16	Q9RRL5	Q9rrl5 deinococcus
16	49.5	38.1	192	15	Q994Q2	Q994q2 human immun
17	49.5	38.1	192	15	Q9QSR2	Q9qsr2 human immun
18	49.5	38.1	603	16	Q8UAZ8	Q8uaz8 agrobacteri
19	49	37.7	90	15	Q9QQA2	Q9qqa2 human immun
20	49	37.7	90	15	Q9QQA0	Q9qqa0 human immun
21	49	37.7	90	15	Q9QQA1	Q9qqa1 human immun
22	49	37.7	90	15	Q9QQ96	Q9qq96 human immun
23	49	37.7	277	16	Q87RD3	Q87rd3 vibrio para
24	49	37.7	331	16	Q7WEG7	Q7weg7 bordetella
25	49	37.7	331	16	Q7W350	Q7w350 bordetella
26	49	37.7	331	16	Q7W032	Q7w032 bordetella
27	49	37.7	363	13	Q9DDH0	Q9ddh0 neoceratodu
28	49	37.7	711	15	Q7SVK4	Q7svk4 human immun
29	49	37.7	711	15	Q7SVK3	Q7svk3 human immun
30	49	37.7	713	15	Q7SVK2	Q7svk2 human immun
31	49	37.7	753	16	Q89T31	Q89t31 bradyrhizob
32	49	37.7	814	16	Q8PHY9	Q8phy9 xanthomonas
33	49	37.7	814	16	Q8P6N5	Q8p6n5 xanthomonas
34	49	37.7	4823	13	O93321	O93321 fugu rubrip
35	48.5	37.3	192	15	Q90CX5	Q90cx5 human immun
36	48.5	37.3	192	15	Q9IW51	Q9iw51 human immun
37	48.5	37.3	192	15	Q8UTJ5	Q8utj5 human immun
38	48.5	37.3	192	15	O70887	O70887 human immun
39	48.5	37.3	195	3	Q9USC7	Q9usc7 schizosacch
40	48	36.9	151	16	Q7U7G9	Q7u7g9 synechococc
41	48	36.9	192	15	Q8UTQ8	Q8utq8 human immun
42	48	36.9	192	15	Q90D29	Q90d29 human immun
43	48	36.9	271	16	Q87F13	Q87f13 xylella fas
44	48	36.9	278	16	Q9CMF6	Q9cmf6 pasteurella
45	48	36.9	300	16	Q8EMZ7	Q8emz7 oceanobacil

ALIGNMENTS

RESULT 1
Q8K408
ID Q8K408 PRELIMINARY; PRT; 41 AA.
AC Q8K408;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Truncated macrophage colony stimulating factor.
GN CSF1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEW.tl;
RX MEDLINE=22069908; PubMed=12074592;
RA Dobbins D.E., Sood R., Hashiramoto A., Hansen C.T., Wilder R.L.,
RA Remmers E.F.;
RT "Mutation of macrophage colony stimulating factor (Csf1) causes
RT osteopetrosis in the tl rat.";
RL Biochem. Biophys. Res. Commun. 294:1114-1120(2002).
DR EMBL; AF514357; AAM54137.1; -.
SQ SEQUENCE 41 AA; 4178 MW; 1D342C19BD18AA41 CRC64;

Query Match 51.9%; Score 67.5; DB 11; Length 41;
Best Local Similarity 68.0%; Pred. No. 0.017;
Matches 17; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

```
Qy          1 MAGLPAVVGLSPGEQEYHRGGVGVL 25
              | ||||  |||| ||| | ||| |
Db         18 MDGLPAAAGLSPREQECCR-GVGAL 41
```

RESULT 2
Q89Y25
ID Q89Y25 PRELIMINARY; PRT; 242 AA.
AC Q89Y25;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)

```
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Bll0130 protein.
GN   BLL0130.
OS   Bradyrhizobium japonicum.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Bradyrhizobiaceae; Bradyrhizobium.
OX   NCBI_TaxID=375;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=USDA 110;
RX   MEDLINE=22484998; PubMed=12597275;
RA   Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA   Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA   Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA   Tabata S.;
RT   "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT   Bradyrhizobium japonicum USDA110.";
RL   DNA Res. 9:189-197(2002).
DR   EMBL; AP005935; BAC45395.1; -.
DR   InterPro; IPR005493; Methyltransf_6.
DR   Pfam; PF03737; Methyltransf_6; 1.
KW   Complete proteome.
SQ   SEQUENCE   242 AA;  25216 MW;  588F9C0B9396414B CRC64;

  Query Match             43.1%;  Score 56;  DB 16;  Length 242;
  Best Local Similarity   57.1%;  Pred. No. 5.5;
  Matches   12;  Conservative    1;  Mismatches    4;  Indels    4;  Gaps    1;

Qy          1 MAGLPAVVGLSPGEQEYHRGG 21
              :| ||  ||| |    |||||
Db         63 LARLPGAVGLKP----YHRGG 79


RESULT 3
Q87YM5
ID   Q87YM5      PRELIMINARY;      PRT;   330 AA.
AC   Q87YM5;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Conserved hypothetical protein.
GN   PSPTO3770.
OS   Pseudomonas syringae (pv. tomato).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=323;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DC3000;
RA   Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA   Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA   Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA   Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
RA   White O., Fraser C., Collmer A.;
RT   "Complete sequence of Pseudomonas syringae.";
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE016869; AAO57239.1; -.
DR   TIGR; PSPTO3770; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   330 AA;  36369 MW;  99876752E292FA71 CRC64;

  Query Match             41.5%;  Score 54;  DB 16;  Length 330;
  Best Local Similarity   43.5%;  Pred. No. 15;
  Matches   10;  Conservative    4;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          3 GLPAVVGLSPGEQEYHRGGVGVL 25
              | | : ||:| |   :  ||:| |
Db        308 GQPLINGLAPSEAVFQPGGIGKL 330


RESULT 4
Q8PE55
ID   Q8PE55      PRELIMINARY;      PRT;   150 AA.
AC   Q8PE55;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Deoxycytidylate deaminase.
GN   XCC0126.
OS   Xanthomonas campestris (pv. campestris).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xanthomonas.
OX   NCBI_TaxID=340;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33913 / NCPPB 528;
RX   MEDLINE=22022145; PubMed=12024217;
RA   da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA   Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA   Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA   Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA   Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA   Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA   Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA   Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA   Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA   Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA   Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA   Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA   Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA   Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA   Setubal J.C., Kitajima J.P.;
RT   "Comparison of the genomes of two Xanthomonas pathogens with differing
RT   host specificities.";
RL   Nature 417:459-463(2002).
DR   EMBL; AE012107; AAM39445.1; -.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   InterPro; IPR002125; dCMP/cyt_deam.
DR   Pfam; PF00383; dCMP_cyt_deam; 1.
KW   Complete proteome.
SQ   SEQUENCE   150 AA;  16201 MW;  63D9C17D44DC8B43 CRC64;

  Query Match             40.0%;  Score 52;  DB 16;  Length 150;
  Best Local Similarity   58.8%;  Pred. No. 12;
  Matches   10;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 AGLPAVVGLSPGEQEYH 18
              ||:  ||  |:||| | |
Db        104 AGIKRVVALAPGESESH 120


RESULT 5
Q9YBR0
ID   Q9YBR0      PRELIMINARY;      PRT;   426 AA.
AC   Q9YBR0;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein APE1539.
GN   APE1539.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC   Desulfurococcaceae; Aeropyrum.
OX   NCBI_TaxID=56636;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE=99310339; PubMed=10382966;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA   Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA   Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA   Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA   Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
```

```
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000061; BAA80538.1; -.
DR   PIR; D72635; D72635.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   426 AA;  48461 MW;  6D3D3469392D4DF0 CRC64;

  Query Match             39.2%;  Score 51;  DB 17;  Length 426;
  Best Local Similarity   45.5%;  Pred. No. 54;
  Matches   10;  Conservative    4;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          3 GLPAVVGLSPGEQEYHRGGVGV 24
              ||  ::|: || :|  | || |
Db        278 GLADILGIEPGRREERREGVRV 299


RESULT 6
Q8K2L1
ID   Q8K2L1      PRELIMINARY;      PRT;   383 AA.
AC   Q8K2L1;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Procollagen, type IX, alpha 3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC030945; AAH30945.1; -.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   Pfam; PF01391; Collagen; 6.
DR   ProDom; PD000007; Clg_helix; 3.
KW   Collagen.
SQ   SEQUENCE   383 AA;  35980 MW;  B5F5AF37FF3192CE CRC64;

  Query Match             38.8%;  Score 50.5;  DB 11;  Length 383;
  Best Local Similarity   52.0%;  Pred. No. 57;
  Matches   13;  Conservative    3;  Mismatches    6;  Indels    3;  Gaps    2;

Qy          3 GLPAVVGLSPGE--QEYHRGGVGVL 25
              |:|  ||: |||  :  ||| || |
Db         66 GIPGDVGV-PGERGEAGHRGSVGAL 89


RESULT 7
Q8FVA5
ID   Q8FVA5      PRELIMINARY;      PRT;   601 AA.
AC   Q8FVA5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Dehydratase, IlvD/Edd family.
GN   BRA0940.
OS   Brucella suis.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Brucellaceae; Brucella.
OX   NCBI_TaxID=29461;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1330 / Biovar 1;
RX   MEDLINE=22247741; PubMed=12271122;
RA   Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA   Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA   Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA   Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA   Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA   Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
RT   "The Brucella suis genome reveals fundamental similarities between
RT   animal and plant pathogens and symbionts.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
DR   EMBL; AE014587; AAN34112.1; -.
DR   TIGR; BRA0940; -.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR000581; ILVD_EDD_family.
DR   Pfam; PF00920; ILVD_EDD; 1.
DR   ProDom; PD002691; ILVD_EDD_family; 1.
DR   PROSITE; PS00886; ILVD_EDD_1; 1.
KW   Complete proteome.
SQ   SEQUENCE   601 AA;  65149 MW;  7746F5B27BD64A86 CRC64;

  Query Match             38.8%;  Score 50.5;  DB 16;  Length 601;
  Best Local Similarity   45.8%;  Pred. No. 94;
  Matches   11;  Conservative    2;  Mismatches    6;  Indels    5;  Gaps    1;

Qy          4 LPAVVGLSP-----GEQEYHRGGV 22
              :| :| | |     ||  || |||
Db        310 IPLLVNLQPAGEYLGEDYYHAGGV 333


RESULT 8
Q8YD21
ID   Q8YD21      PRELIMINARY;      PRT;   615 AA.
AC   Q8YD21;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Galactonate dehydratase (EC 4.2.1.6).
GN   BMEII0356.
OS   Brucella melitensis.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Brucellaceae; Brucella.
OX   NCBI_TaxID=29459;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=16M / ATCC 23456 / Biotype 1;
RX   MEDLINE=20020109; PubMed=11756688;
RA   DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujer C., Los T.,
RA   Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA   Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
RA   Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA   Haselkorn R., Kyrpides N., Overbeek R.;
RT   "The genome sequence of the facultative intracellular pathogen
RT   Brucella melitensis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR   EMBL; AE009673; AAL53598.1; -.
DR   PIR; AC3554; AC3554.
DR   GO; GO:0008869; F:galactonate dehydratase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR000581; ILVD_EDD_family.
DR   Pfam; PF00920; ILVD_EDD; 1.
DR   ProDom; PD002691; ILVD_EDD_family; 1.
DR   PROSITE; PS00886; ILVD_EDD_1; 1.
KW   Lyase; Complete proteome.
SQ   SEQUENCE   615 AA;  66750 MW;  F7DC4B656C101438 CRC64;

  Query Match             38.8%;  Score 50.5;  DB 16;  Length 615;
  Best Local Similarity   45.8%;  Pred. No. 96;
  Matches   11;  Conservative    2;  Mismatches    6;  Indels    5;  Gaps    1;

Qy          4 LPAVVGLSP-----GEQEYHRGGV 22
              :| :| | |     ||  || |||
Db        324 IPLLVNLQPAGEYLGEDYYHAGGV 347


RESULT 9
Q8VHY3
ID   Q8VHY3      PRELIMINARY;      PRT;   662 AA.
AC   Q8VHY3;
```

```
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Alpha-3 type IX collagen (Fragment).
GN   COL9A3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Swiss white; TISSUE=Urogenital ridge;
RA   McClive P.J., Bell K.M., Sinclair A.H.;
RT   "Type II and type IX collagen genes are expressed male-specifically in
RT   the developing mouse gonad.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF349718; AAL56219.1; -.
DR   MGD; MGI:894686; Col9a3.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   Pfam; PF01391; Collagen; 11.
DR   ProDom; PD000007; Clg_helix; 3.
KW   Collagen.
FT   NON_TER       1      1
SQ   SEQUENCE   662 AA;  61726 MW;  9261FFBABF1AC4B7 CRC64;

  Query Match             38.8%;  Score 50.5;  DB 11;  Length 662;
  Best Local Similarity   52.0%;  Pred. No. 1e+02;
  Matches   13;  Conservative    3;  Mismatches    6;  Indels    3;  Gaps    2;

Qy          3 GLPAVVGLSPGE--QEYHRGGVGVL 25
              |:|  ||: |||  :  ||| || |
Db        345 GIPGDVGV-PGERGEAGHRGSVGAL 368


RESULT 10
Q8BSX1
ID   Q8BSX1      PRELIMINARY;      PRT;   680 AA.
AC   Q8BSX1;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Procollagen.
GN   COL9A3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Xiphoid cartilage;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK030350; BAC26915.1; -.
DR   MGD; MGI:894686; Col9a3.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   Pfam; PF01391; Collagen; 11.
DR   ProDom; PD000007; Clg_helix; 2.
SQ   SEQUENCE   680 AA;  63501 MW;  1939E44571032AE4 CRC64;

  Query Match             38.8%;  Score 50.5;  DB 11;  Length 680;
  Best Local Similarity   52.0%;  Pred. No. 1.1e+02;
  Matches   13;  Conservative    3;  Mismatches    6;  Indels    3;  Gaps    2;

Qy          3 GLPAVVGLSPGE--QEYHRGGVGVL 25
              |:|  ||: |||  :  ||| || |
Db        363 GIPGDVGV-PGERGEAGHRGSVGAL 386


RESULT 11
Q8L525
ID   Q8L525      PRELIMINARY;      PRT;   758 AA.
AC   Q8L525;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   B1103C09.11 protein (P0451D05.19 protein).
GN   B1103C09.11 OR P0451D05.19.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, BAC
RT   clone:B1103C09.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC
RT   clone:P0451D05.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP003333; BAB91821.1; -.
DR   EMBL; AP003253; BAB92317.1; -.
DR   Gramene; Q8L525; -.
SQ   SEQUENCE   758 AA;  81761 MW;  2A95E763198FE12C CRC64;

  Query Match             38.8%;  Score 50.5;  DB 10;  Length 758;
  Best Local Similarity   32.4%;  Pred. No. 1.2e+02;
  Matches   12;  Conservative    3;  Mismatches    7;  Indels   15;  Gaps    1;

Qy          3 GLPAVVGLSPGEQEY---------------HRGGVGV 24
              |    :||: ||| :                | |||||
Db        552 GAAGAIGLAQGEQNFPGTPALLPVMQFGSQHPGGVGV 588


RESULT 12
Q9KQK6
ID   Q9KQK6      PRELIMINARY;      PRT;   277 AA.
AC   Q9KQK6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Formyltetrahydrofolate deformylase.
GN   VC1992.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=El Tor N16961 / Serotype O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004274; AAF95140.1; -.
DR   PIR; F82130; F82130.
DR   HSSP; P08179; 2GAR.
```

```
DR   TIGR; VC1992; -.
DR   GO; GO:0016597; F:amino acid binding; IEA.
DR   GO; GO:0008864; F:formyltetrahydrofolate deformylase activity; IEA.
DR   GO; GO:0016742; F:hydroxymethyl-, formyl- and related transfe. . .; IEA.
DR   GO; GO:0006189; P:'de novo' IMP biosynthesis; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   InterPro; IPR002912; ACT.
DR   InterPro; IPR002376; formyl_transf.
DR   InterPro; IPR004810; PurU.
DR   Pfam; PF01842; ACT; 1.
DR   Pfam; PF00551; formyl_transf; 1.
DR   PRINTS; PR01575; FFH4HYDRLASE.
DR   TIGRFAMs; TIGR00655; PurU; 1.
KW   Complete proteome.
SQ   SEQUENCE   277 AA;  31373 MW;  A703491654753DC6 CRC64;

  Query Match           38.5%;  Score 50;  DB 16;  Length 277;
  Best Local Similarity  52.9%;  Pred. No. 47;
  Matches    9;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :|  | :| | ||
Db        190 LPAFIGAKPYQQAYERG 206


RESULT 13
Q7VFI5
ID   Q7VFI5      PRELIMINARY;      PRT;   284 AA.
AC   Q7VFI5;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Formyltetrahydrofolate deformylase PurU (EC 3.5.1.10).
GN   PURU OR HH1691.
OS   Helicobacter hepaticus.
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=32025;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 51449 / 3B1;
RX   MEDLINE=22709201; PubMed=12810954;
RA   Suerbaum S., Josenhans C., Sterzenbach T., Drescher B., Brandt P.,
RA   Bell M., Droege M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,
RA   Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,
RA   Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.;
RT   "The complete genome sequence of the carcinogenic bacterium
RT   Helicobacter hepaticus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906(2003).
DR   EMBL; AE017149; AAP78288.1; -.
KW   Hydrolase; Complete proteome.
SQ   SEQUENCE   284 AA;  32660 MW;  AAED2DC5086236E3 CRC64;

  Query Match           38.5%;  Score 50;  DB 16;  Length 284;
  Best Local Similarity  52.9%;  Pred. No. 49;
  Matches    9;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :| :| :| | ||
Db        197 LPAFIGANPYKQAYERG 213


RESULT 14
Q82RR7
ID   Q82RR7      PRELIMINARY;      PRT;   335 AA.
AC   Q82RR7;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative terpene cyclase.
GN   TPC1 OR SAV76.
OS   Streptomyces avermitilis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX   MEDLINE=22608306; PubMed=12692562;
RA   Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA   Sakaki Y., Hattori M., Omura S.;
RT   "Complete genome sequence and comparative analysis of the industrial
RT   microorganism Streptomyces avermitilis.";
RL   Nat. Biotechnol. 21:526-531(2003).
DR   EMBL; AP005021; BAC67785.1; -.
DR   InterPro; IPR008949; Terpenoid_synth.
KW   Complete proteome.
SQ   SEQUENCE   335 AA;  36480 MW;  49B8477E2D52666F CRC64;

  Query Match           38.5%;  Score 50;  DB 16;  Length 335;
  Best Local Similarity  47.6%;  Pred. No. 58;
  Matches   10;  Conservative    3;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          3 GLPAVVGLSPGEQEYHRGGVG 23
              |||| |:||| :   |  :|
Db          8 GLPAPAGISPGLEATRRHNLG 28


RESULT 15
Q9RRL5
ID   Q9RRL5      PRELIMINARY;      PRT;   381 AA.
AC   Q9RRL5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cytochrome P450.
GN   DR2473.
OS   Deinococcus radiodurans.
OC   Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AE002076; AAF12016.1; -.
DR   PIR; F75270; F75270.
DR   TIGR; DR2473; -.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
```

KW Heme; Monooxygenase; Oxidoreductase; Complete proteome.
SQ SEQUENCE 381 AA; 41940 MW; F191EA69F1797B53 CRC64;

Query Match 38.5%; Score 50; DB 16; Length 381;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          3 GLPAVVGLSP 12
              ||||||||||
Db         51 GLPAVVGLSP 60
```

Search completed: May 7, 2004, 12:37:50
Job time : 48.9167 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:19:40 ; Search time 69.75 Seconds
(without alignments)
101.272 Million cell updates/sec

Title: US-09-786-214A-5
Perfect score: 130
Sequence: 1 MAGLPAVVGLSPGEQEYHRGGVGVL 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	130	100.0	25	3	AAY84264	Aay84264 Peptide o
2	106	81.5	20	3	AAY84265	Aay84265 Truncated
3	75	57.7	15	3	AAY84269	Aay84269 Peptide d
4	72	55.4	14	3	AAY84266	Aay84266 Peptide d
5	68	52.3	13	3	AAY84267	Aay84267 Peptide d
6	65	50.0	13	3	AAY84268	Aay84268 Peptide d
7	53	40.8	394	4	AAU43816	Aau43816 Propionib
8	53	40.8	394	6	ABM40335	Abm40335 Propionib
9	52	40.0	187	6	ABU21296	Abu21296 Protein e
10	51.5	39.6	191	3	AAB69296	Aab69296 HIV-1 non
11	50	38.5	277	6	ABU49441	Abu49441 Protein e
12	49.5	38.1	192	3	AAB69298	Aab69298 HIV-1 non
13	49	37.7	278	6	ABU30626	Abu30626 Protein e
14	48.5	37.3	192	3	AAB69290	Aab69290 HIV-1 non
15	48	36.9	278	6	ABU39195	Abu39195 Protein e
16	48	36.9	287	6	ABU17217	Abu17217 Protein e
17	48	36.9	303	6	ADA36829	Ada36829 Acinetoba
18	47.5	36.5	114	4	AAE03295	Aae03295 Human gen
19	47.5	36.5	114	5	ABG64478	Abg64478 Human alb
20	47	36.2	151	5	ABU50765	Abu50765 Helicobac
21	47	36.2	155	5	ABU51750	Abu51750 Helicobac
22	47	36.2	238	6	ABU27483	Abu27483 Protein e
23	47	36.2	278	6	ABU44771	Abu44771 Protein e
24	47	36.2	280	6	ABU48054	Abu48054 Protein e
25	47	36.2	280	6	ABU31449	Abu31449 Protein e
26	47	36.2	280	6	ABU15054	Abu15054 Protein e
27	47	36.2	280	6	ABU47098	Abu47098 Protein e
28	47	36.2	282	6	ABU41187	Abu41187 Protein e
29	47	36.2	283	6	ABM69037	Abm69037 Photorhab
30	47	36.2	293	2	AAW98485	Aaw98485 H. pylori
31	47	36.2	293	6	ABU31073	Abu31073 Protein e
32	46.5	35.8	656	6	ABU27276	Abu27276 Protein e
33	46	35.4	46	3	AAY58745	Aay58745 Somatosta
34	46	35.4	143	5	AAO21677	Aao21677 Human sec
35	46	35.4	143	7	ADB64633	Adb64633 Human pro
36	46	35.4	184	5	ABP26551	Abp26551 Streptoco
37	46	35.4	234	4	AAB36208	Aab36208 Human imm
38	46	35.4	274	6	ABU26498	Abu26498 Protein e
39	46	35.4	287	7	ADB80057	Adb80057 Mycobacte
40	46	35.4	300	2	AAY16108	Aay16108 A formate
41	46	35.4	418	5	ABB04480	Abb04480 Arthrobac
42	46	35.4	450	5	ABB04479	Abb04479 Arthrobac
43	46	35.4	692	6	ABR40877	Abr40877 Oryza sat
44	46	35.4	921	2	AAY08304	Aay08304 Human col
45	46	35.4	1070	7	ADE08073	Ade08073 Novel pro

ALIGNMENTS

RESULT 1
AAY84264
ID AAY84264 standard; peptide; 25 AA.
XX
AC AAY84264;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide of alternate reading frame of macrophage colony stimulating gene.
XX
KW Renal cell carcinoma; antigen; cytotoxic T lymphocyte;
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
DR N-PSDB; AAZ99672.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Claim 1; Page 64; 74pp; English.
XX
CC The present sequence represents a tumour rejection antigen precursor, and
CC is encoded by an alternative open reading frame (ORF) of human macrophage
CC colony stimulating gene. Peptides derived from the alternative ORF of
CC macrophage-colony stimulating factor, when presented by an antigen
CC presenting cell having a human leukocyte antigen (HLA) class I molecule,
CC effectively induce the activation and proliferation of CD8+ cytotoxic T
CC lymphocytes. Polypeptide and nucleic acids derived from the alternate ORF
CC of macrophage-colony stimulating factor are useful for enriching
CC selectively a population of T lymphocytes with CD8+ T lymphocytes. They
CC are also useful for diagnosing a disorder characterized by expression of

```
CC   the polypeptide, and for identifying functional variants and mimetics
XX
SQ   Sequence 25 AA;

  Query Match             100.0%;  Score 130;  DB 3;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-11;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAGLPAVVGLSPGEQEYHRGGVGVL 25
              |||||||||||||||||||||||||
Db          1 MAGLPAVVGLSPGEQEYHRGGVGVL 25


RESULT 2
AAY84265
ID   AAY84265 standard; peptide; 20 AA.
XX
AC   AAY84265;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Truncated macrophage colony stimulating factor tumour antigen.
XX
KW   tumour rejection antigen; macrophage colony stimulating gene;
KW   macrophage-colony stimulating factor; antigen presenting cell;
KW   human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS   Homo sapiens.
XX
PN   WO200013699-A1.
XX
PD   16-MAR-2000.
XX
PF   03-SEP-1999;   99WO-US020344.
XX
PR   04-SEP-1998;   98US-0099077P.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Probst-Kepper M,  Van Den Eynde B,  Boon-Falleur T;
XX
DR   WPI; 2000-256859/22.
DR   N-PSDB; AAZ99675.
XX
PT   Isolated polypeptide used to treat subjects having a disorder
PT   characterized by expression of alternative open reading frame macrophage-
PT   colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS   Claim 3; Page 64; 74pp; English.
XX
CC   The present sequence represents a truncated tumour rejection antigen
CC   precursor, and is encoded by a truncated alternative open reading frame
CC   (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC   the alternative ORF of macrophage-colony stimulating factor, when
CC   presented by an antigen presenting cell having a human leukocyte antigen
CC   (HLA) class I molecule, effectively induce the activation and
CC   proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC   acids derived from the alternate ORF of macrophage-colony stimulating
CC   factor are useful for enriching selectively a population of T lymphocytes
CC   with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC   characterized by expression of the polypeptide, and for identifying
CC   functional variants and mimetics
XX
SQ   Sequence 20 AA;

  Query Match             81.5%;  Score 106;  DB 3;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-08;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAGLPAVVGLSPGEQEYHRG 20
              ||||||||||||||||||||
Db          1 MAGLPAVVGLSPGEQEYHRG 20


RESULT 3
AAY84269
ID   AAY84269 standard; peptide; 15 AA.
XX
AC   AAY84269;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW   tumour rejection antigen; macrophage colony stimulating gene;
KW   macrophage-colony stimulating factor; antigen presenting cell;
KW   human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO200013699-A1.
XX
PD   16-MAR-2000.
XX
PF   03-SEP-1999;   99WO-US020344.
XX
PR   04-SEP-1998;   98US-0099077P.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Probst-Kepper M,  Van Den Eynde B,  Boon-Falleur T;
XX
DR   WPI; 2000-256859/22.
XX
PT   Isolated polypeptide used to treat subjects having a disorder
PT   characterized by expression of alternative open reading frame macrophage-
PT   colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS   Example 2; Page 40; 74pp; English.
XX
CC   The present sequence represents a peptide which is derived from a tumour
CC   rejection antigen precursor encoded by an alternative open reading frame
CC   (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC   the alternative ORF of macrophage-colony stimulating factor, when
CC   presented by an antigen presenting cell having a human leukocyte antigen
CC   (HLA) class I molecule, effectively induce the activation and
CC   proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC   acids derived from the alternate ORF of macrophage-colony stimulating
CC   factor are useful for enriching selectively a population of T lymphocytes
CC   with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC   characterized by expression of the polypeptide, and for identifying
CC   functional variants and mimetics
XX
SQ   Sequence 15 AA;

  Query Match             57.7%;  Score 75;  DB 3;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.00087;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 AGLPAVVGLSPGEQE 16
              |||||||||||||||
Db          1 AGLPAVVGLSPGEQE 15


RESULT 4
AAY84266
ID   AAY84266 standard; peptide; 14 AA.
XX
AC   AAY84266;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW   tumour rejection antigen; macrophage colony stimulating gene;
```

KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Claim 2; Page 39; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics
XX
SQ Sequence 14 AA;

Query Match 55.4%; Score 72; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0022;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 LPAVVGLSPGEQEY 17
||||||||||||||
Db 1 LPAVVGLSPGEQEY 14

RESULT 5
AAY84267
ID AAY84267 standard; peptide; 13 AA.
XX
AC AAY84267;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Example 2; Page 40; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics
XX
SQ Sequence 13 AA;

Query Match 52.3%; Score 68; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.0074;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 PAVVGLSPGEQEY 17
|||||||||||||
Db 1 PAVVGLSPGEQEY 13

RESULT 6
AAY84268
ID AAY84268 standard; peptide; 13 AA.
XX
AC AAY84268;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX

PS Example 2; Page 40; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics
XX
SQ Sequence 13 AA;

Query Match 50.0%; Score 65; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 LPAVVGLSPGEQE 16
|||||||||||||
Db 1 LPAVVGLSPGEQE 13

RESULT 7
AAU43816
ID AAU43816 standard; protein; 394 AA.
XX
AC AAU43816;
XX
DT 13-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #4712.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US012865.
XX
PR 21-APR-2000; 2000US-0199047P.
PR 02-JUN-2000; 2000US-0208841P.
PR 07-JUL-2000; 2000US-0216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59521.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
PS Claim 6; SEQ ID NO 5011; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 394 AA;

Query Match 40.8%; Score 53; DB 4; Length 394;
Best Local Similarity 45.0%; Pred. No. 39;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 5 PAVVGLSPGEQEYHRGGVGV 24
| |:| :|| ::: | | ||
Db 255 PVVLGTAPGGEQHDRHGTGV 274

RESULT 8
ABM40335
ID ABM40335 standard; protein; 394 AA.
XX
AC ABM40335;
XX
DT 20-OCT-2003 (first entry)
XX
DE Propionibacterium acnes immunogenic polypeptide #5011.
XX
KW Acne vulgaris; antiseborrhoeic; dermatological; antibacterial;
KW immunostimulant; immune response; vaccine; immunogenic.
XX
OS Propionibacterium acnes.
XX
PN WO2003033515-A1.
XX
PD 24-APR-2003.
XX
PF 11-OCT-2002; 2002WO-US032727.
XX
PR 15-OCT-2001; 2001US-00978825.
XX
PA (CORI-) CORIXA CORP.
XX
PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
PI Barth B, Vallieve-Douglass J;
XX
DR WPI; 2003-381789/36.
DR N-PSDB; ACF64450.
XX
PT New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.
XX
PS Claim 6; SEQ ID NO 5011; 1481pp; English.
XX
CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared

```
CC   via this method; a vaccine composition (comprising P. acnes polypeptides,
CC   polynucleotides, antibodies, fusion proteins, T cell populations, or
CC   antigen-presenting cells that express the polypeptide); a method and kit
CC   for detecting or determining the presence or absence of P. acnes in a
CC   patient; and a method for inhibiting the development of P. acnes in a
CC   patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC   proteins, T cell populations or antigen-presenting cells that express the
CC   polypeptides are useful for diagnosing, preventing or treating acne
CC   vulgaris, or for stimulating an immune response specific for a P. acnes
CC   protein. The polynucleotides can also be used as probes or primers for
CC   nucleic acid hybridisation. The vaccine composition is useful for the
CC   stimulation of an immune response against P. acnes, or for treating acne,
CC   and the kit is useful for performing a diagnostic assay. The present
CC   sequence represents a specifically claimed P. acnes polypeptide which is
CC   thought to contain an immunogenic region. Note: The sequence data for
CC   this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 394 AA;

  Query Match             40.8%;  Score 53;  DB 6;  Length 394;
  Best Local Similarity   45.0%;  Pred. No. 39;
  Matches    9;  Conservative    5;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          5 PAVVGLSPGEQEYHRGGVGV 24
              | |:| :|| ::: | | ||
Db        255 PVVLGTAPGGEQHDRHGTGV 274


RESULT 9
ABU21296
ID   ABU21296 standard; protein; 187 AA.
XX
AC   ABU21296;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #6823.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Burkholderia fungorum.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA25166.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 49220; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 187 AA;

  Query Match             40.0%;  Score 52;  DB 6;  Length 187;
  Best Local Similarity   47.8%;  Pred. No. 24;
  Matches   11;  Conservative    2;  Mismatches    6;  Indels    4;  Gaps    1;

Qy          3 GLPA----VVGLSPGEQEYHRGG 21
              |||     |||::||  | |  |
Db         48 GLPVVLVNVVGMAPGRTEQHHAG 70


RESULT 10
AAB69296
ID   AAB69296 standard; protein; 191 AA.
XX
AC   AAB69296;
XX
DT   12-SEP-2003  (revised)
DT   20-APR-2001  (first entry)
XX
DE   HIV-1 non-subtype B clone 94CY032-3 vif protein.
XX
KW   HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW   vif; vpr; tat; rev; nef; vaccine.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO200026416-A1.
XX
PD   11-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US024837.
XX
PR   02-NOV-1998;   98US-00184418.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Hahn BH,  Shaw GM,  Gao F;
XX
DR   WPI; 2000-365651/31.
XX
PT   Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT   type 1 useful for detecting and treating AIDS comprises a specific
PT   nucleotide sequence.
```

XX
PS Claim 41; Fig 16; 131pp; English.
XX
CC The present in invention provides the protein and coding sequences for a
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC isolates. The sequences shown include the near full-length coding
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC rev and nef proteins. These can be used to detect the presence of HIV-1
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC These antibodies can be used in vaccines to prevent and treat HIV
CC infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ Sequence 191 AA;

Query Match 39.6%; Score 51.5; DB 3; Length 191;
Best Local Similarity 58.8%; Pred. No. 29;
Matches 10; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

Qy 9 GLSPGEQEYHRG-GVGV 24
|| |||| :: | | || :
Db 70 GLQPGEQDWHLGHGVSI 86

RESULT 11
ABU49441
ID ABU49441 standard; protein; 277 AA.
XX
AC ABU49441;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #34968.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Vibrio cholerae.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA53311.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 77365; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 277 AA;

Query Match 38.5%; Score 50; DB 6; Length 277;
Best Local Similarity 52.9%; Pred. No. 71;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 4 LPAVVGLSPGEQEYHRG 20
||| :| | :| | ||
Db 190 LPAFIGAKPYQQAYERG 206

RESULT 12
AAB69298
ID AAB69298 standard; protein; 192 AA.
XX
AC AAB69298;
XX
DT 12-SEP-2003 (revised)
DT 20-APR-2001 (first entry)
XX
DE HIV-1 non-subtype B clone 96ZM751-3 vif protein.
XX
KW HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW vif; vpr; tat; rev; nef; vaccine.
XX
OS Human immunodeficiency virus 1.
XX
PN WO200026416-A1.
XX
PD 11-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US024837.
XX
PR 02-NOV-1998; 98US-00184418.
XX
PA (UABR-) UAB RES FOUND.
XX
PI Hahn BH, Shaw GM, Gao F;
XX
DR WPI; 2000-365651/31.
XX
PT Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT type 1 useful for detecting and treating AIDS comprises a specific
PT nucleotide sequence.
XX
PS Claim 41; Fig 16; 131pp; English.
XX
CC The present in invention provides the protein and coding sequences for a
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC isolates. The sequences shown include the near full-length coding
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,

```
CC   rev and nef proteins. These can be used to detect the presence of HIV-1
CC   in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC   These antibodies can be used in vaccines to prevent and treat HIV
CC   infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 192 AA;

  Query Match             38.1%;  Score 49.5;  DB 3;  Length 192;
  Best Local Similarity   58.8%;  Pred. No. 56;
  Matches   10;  Conservative    3;  Mismatches    3;  Indels    1;  Gaps    1;

Qy          9 GLSPGEQEYHRG-GVGV 24
              || |||:|:| | || :
Db         71 GLHPGEREWHLGHGVSI 87


RESULT 13
ABU30626
ID   ABU30626 standard; protein; 278 AA.
XX
AC   ABU30626;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #16153.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Haemophilus influenzae.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA34496.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 58550; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 278 AA;

  Query Match             37.7%;  Score 49;  DB 6;  Length 278;
  Best Local Similarity   52.9%;  Pred. No. 98;
  Matches    9;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :|  | :| | ||
Db        191 LPAFIGAKPYQQAYKRG 207


RESULT 14
AAB69290
ID   AAB69290 standard; protein; 192 AA.
XX
AC   AAB69290;
XX
DT   12-SEP-2003  (revised)
DT   20-APR-2001  (first entry)
XX
DE   HIV-1 non-subtype B clone 93BR020-1 vif protein.
XX
KW   HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW   vif; vpr; tat; rev; nef; vaccine.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO200026416-A1.
XX
PD   11-MAY-2000.
XX
PF   25-OCT-1999;   99WO-US024837.
XX
PR   02-NOV-1998;   98US-00184418.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Hahn BH,  Shaw GM,  Gao F;
XX
DR   WPI; 2000-365651/31.
XX
PT   Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT   type 1 useful for detecting and treating AIDS comprises a specific
PT   nucleotide sequence.
XX
PS   Claim 41; Fig 16; 131pp; English.
XX
CC   The present in invention provides the protein and coding sequences for a
CC   number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC   isolates. The sequences shown include the near full-length coding
CC   sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC   rev and nef proteins. These can be used to detect the presence of HIV-1
CC   in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC   These antibodies can be used in vaccines to prevent and treat HIV
CC   infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 192 AA;
```

```
  Query Match             37.3%;  Score 48.5;  DB 3;  Length 192;
  Best Local Similarity   58.8%;  Pred. No. 78;
  Matches   10;  Conservative    3;  Mismatches    3;  Indels    1;  Gaps    1;

Qy          9 GLSPGEQEYHRG-GVGV 24
              || |||:|:| | || :
Db         71 GLLPGEREWHLGQGVSI 87


RESULT 15
ABU39195
ID   ABU39195 standard; protein; 278 AA.
XX
AC   ABU39195;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #24722.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Pasteurella multocida.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA43065.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 67119; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 278 AA;

  Query Match             36.9%;  Score 48;  DB 6;  Length 278;
  Best Local Similarity   52.9%;  Pred. No. 1.4e+02;
  Matches    9;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :|  |  | | ||
Db        191 LPAFIGAKPYHQAYERG 207


Search completed: May  7, 2004, 12:33:42
Job time : 71.75 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:20:30 ; Search time 7.2 Seconds
(without alignments)
144.639 Million cell updates/sec

Title: US-09-786-214A-9
Perfect score: 106
Sequence: 1 MAGLPAVVGLSPGEQEYHRG 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	49	46.2	278	1	PURU_HAEIN	Q03432 haemophilus
2	48	45.3	780	1	PPSA_DEIRA	O83026 deinococcus
3	47	44.3	280	1	PURU_ECOL6	P38480 escherichia
4	47	44.3	280	1	PURU_ECOLI	P37051 escherichia
5	46	43.4	508	1	COBI_MYCTU	Q10677 mycobacteri
6	44	41.5	1172	1	CNA2_MOUSE	Q80xc6 mus musculu
7	44	41.5	3664	1	MINT_HUMAN	Q96t58 homo sapien
8	43	40.6	211	1	COBL_METJA	Q58917 methanococc
9	43	40.6	326	1	MER_METTI	Q9uxp0 methanolobu
10	43	40.6	390	1	COBL_MYCTU	Q10671 mycobacteri
11	43	40.6	505	1	YT6J_CAEEL	P90838 caenorhabdi
12	43	40.6	516	1	C4AD_DROME	Q9v4t3 drosophila
13	43	40.6	544	1	CH60_PROAC	Q9k2u4 propionibac
14	43	40.6	628	1	HNFA_MOUSE	P22361 mus musculu
15	43	40.6	628	1	HNFA_RAT	P15257 rattus norv
16	43	40.6	631	1	HNFA_HUMAN	P20823 homo sapien
17	43	40.6	633	1	PLB5_SCHPO	Q9y7n6 schizosacch
18	43	40.6	658	1	VG18_BPT4	P13332 bacteriopha
19	43	40.6	677	1	SG1_HUMAN	P05060 homo sapien
20	43	40.6	983	1	RRPO_SJNNV	Q9qaz8 striped jac
21	42	39.6	400	1	TRUD_METKA	Q8txj7 methanopyru
22	42	39.6	469	1	GLYC_YEAST	P37291 saccharomyc
23	42	39.6	808	1	PLSB_VIBPA	Q87kn0 vibrio para
24	42	39.6	876	1	TOP1_VIBCH	Q9krb2 vibrio chol
25	42	39.6	1065	1	RPOB_MARPO	P06272 marchantia
26	42	39.6	1712	1	CA24_HUMAN	P08572 homo sapien
27	41	38.7	156	1	RUVX_CAUCR	Q9a5k8 caulobacter
28	41	38.7	213	1	MDCG_XANAC	Q8ppw9 xanthomonas
29	41	38.7	243	1	SUMT_SYNP7	P42451 synechococc
30	41	38.7	479	1	PTSB_VIBAL	P22825 vibrio algi
31	41	38.7	481	1	GYRB_FLEJA	Q9fax2 flexibacter
32	41	38.7	520	1	AT15_YEAST	P25641 saccharomyc
33	41	38.7	579	1	NH71_CAEEL	Q9gtd4 caenorhabdi
34	41	38.7	774	1	LOL2_HUMAN	Q9y4k0 homo sapien
35	41	38.7	934	1	ODO1_COXBU	P51056 coxiella bu
36	41	38.7	981	1	RRPO_AHNNV	Q9dic5 atlantic ha
37	40.5	38.2	387	1	OMA2_NEIMC	P18194 neisseria m
38	40.5	38.2	392	1	OMA_NEIMB	Q51240 neisseria m
39	40.5	38.2	393	1	OMA1_NEIMC	P13415 neisseria m
40	40.5	38.2	395	1	OMA_NEIMA	P57041 neisseria m
41	40.5	38.2	811	1	PRIA_RHORU	P05445 rhodospiril
42	40	37.7	157	1	RISB_PYRFU	Q8u418 pyrococcus
43	40	37.7	193	1	VIF_SIVCZ	P17284 chimpanzee
44	40	37.7	316	1	Y273_SYNY3	P73894 synechocyst
45	40	37.7	331	1	KDD2_STRAW	Q82bx4 streptomyce

ALIGNMENTS

RESULT 1
PURU_HAEIN
ID PURU_HAEIN STANDARD; PRT; 278 AA.
AC Q03432;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-FH(4)
DE hydrolase).
GN PURU OR HI1588.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
RN [2]
RP SEQUENCE OF 64-278 FROM N.A.
RC STRAIN=RM 7004 / Serotype B;
RX MEDLINE=93328119; PubMed=8335255;
RA Maskell D.J.;
RT "Cloning and sequencing of the Haemophilus influenzae aroA gene.";
RL Gene 129:155-156(1993).
CC -!- FUNCTION: PRODUCES FORMATE FROM FORMYL-TETRAHYDROFOLATE. PROVIDES
CC THE MAJOR SOURCE OF FORMATE FOR THE PURT-DEPENDENT SYNTHESIS OF
CC 5'-PHOSPHORIBOSYL-N-FORMYLGLYCINAMIDE (FGAR) DURING AEROBIC
CC GROWTH. HAS A ROLE IN REGULATING THE ONE-CARBON POOL
CC (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + H(2)O = formate +
CC tetrahydrofolate.
CC -!- ENZYME REGULATION: Activated by methionine, inhibited by glycine
CC (By similarity).
CC -!- PATHWAY: De novo purine biosynthesis.
CC -!- SUBUNIT: Homohexamer (By similarity).
CC -!- SIMILARITY: SOME, TO GAR TRANSFORMYLASE (PURN).
CC ---

```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32833; AAC23236.1; -.
DR   EMBL; L04686; AAA24942.1; -.
DR   PIR; E64131; E64131.
DR   HSSP; P08179; 1GRC.
DR   TIGR; HI1588; -.
DR   InterPro; IPR002912; ACT.
DR   InterPro; IPR002376; formyl_transf.
DR   InterPro; IPR004810; PurU.
DR   Pfam; PF01842; ACT; 1.
DR   Pfam; PF00551; formyl_transf; 1.
DR   PRINTS; PR01575; FFH4HYDRLASE.
DR   TIGRFAMs; TIGR00655; PurU; 1.
KW   Purine biosynthesis; Hydrolase; One-carbon metabolism;
KW   Complete proteome.
FT   ACT_SITE    223    223       BY SIMILARITY.
FT   CONFLICT    115    117       VIG -> RNR (IN REF. 2).
FT   CONFLICT    138    140       HEN -> PK (IN REF. 2).
FT   CONFLICT    205    205       K -> E (IN REF. 2).
SQ   SEQUENCE   278 AA;  32173 MW;  7F375AB3C422EC4B CRC64;

  Query Match             46.2%;  Score 49;  DB 1;  Length 278;
  Best Local Similarity   52.9%;  Pred. No. 3.6;
  Matches    9;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :|  | :| | ||
Db        191 LPAFIGAKPYQQAYKRG 207


RESULT 2
PPSA_DEIRA
ID   PPSA_DEIRA     STANDARD;      PRT;   780 AA.
AC   O83026;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phosphoenolpyruvate synthase (EC 2.7.9.2) (Pyruvate,water dikinase)
DE   (PEP synthase).
GN   PPSA OR DR1727.
OS   Deinococcus radiodurans.
OC   Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
RN   [2]
RP   SEQUENCE OF 259-780 FROM N.A.
RA   Narumi I., Islam S., Cherdchu K., Kikuchi M., Watanabe H.,
RA   Kitayama S., Yamamoto K.;
RT   "IS8301: the second insertion sequence element from Deinococcus
RT   radiodurans.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ATP + pyruvate + H(2)O = AMP +
CC       phosphoenolpyruvate + phosphate.
CC   -!- PATHWAY: ESSENTIAL STEP IN GLUCONEOGENESIS WHEN PYRUVATE AND
CC       LACTATE ARE USED AS A CARBON SOURCE.
CC   -!- SIMILARITY: Belongs to the PEP-utilizing enzyme family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE002014; AAF11283.1; -.
DR   EMBL; AB016803; BAA32387.1; -.
DR   PIR; D75361; D75361.
DR   PIR; T44369; T44369.
DR   TIGR; DR1727; -.
DR   InterPro; IPR008279; PEP_mobile.
DR   InterPro; IPR006318; PEP_P_trans.
DR   InterPro; IPR006319; PEP_synth.
DR   InterPro; IPR000121; PEP_utilizers.
DR   InterPro; IPR002192; PPDK_N_term.
DR   Pfam; PF00391; PEP-utilizers; 1.
DR   Pfam; PF02896; PEP-utilizers_C; 1.
DR   Pfam; PF01326; PPDK_N_term; 1.
DR   PRINTS; PR01736; PHPHTRNFRASE.
DR   ProDom; PD000940; PEP_utilizers; 1.
DR   TIGRFAMs; TIGR01418; PEP_synth; 1.
DR   PROSITE; PS00370; PEP_ENZYMES_PHOS_SITE; 1.
DR   PROSITE; PS00742; PEP_ENZYMES_2; 1.
KW   Transferase; Kinase; ATP-binding; Phosphorylation; Complete proteome.
FT   ACT_SITE    409    409       TELE-PHOSPHOHISTIDINE INTERMEDIATE
FT                                (BY SIMILARITY).
FT   MOD_RES     409    409       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   780 AA;  84895 MW;  AD555076324ADA47 CRC64;

  Query Match             45.3%;  Score 48;  DB 1;  Length 780;
  Best Local Similarity   50.0%;  Pred. No. 15;
  Matches    9;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          3 GLPAVVGLSPGEQEYHRG 20
              |:|||||     :| | |
Db        418 GIPAVVGTGNATRELHNG 435


RESULT 3
PURU_ECOL6
ID   PURU_ECOL6     STANDARD;      PRT;   280 AA.
AC   P38480;
DT   01-OCT-1994 (Rel. 30, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-FH(4)
DE   hydrolase).
GN   PURU OR C1696 OR SF1232 OR S1318.
OS   Escherichia coli O6, and
OS   Shigella flexneri.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=217992, 623;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX   MEDLINE=22388234; PubMed=12471157;
RA   Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA   Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA   Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA   Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT   "Extensive mosaic structure revealed by the complete genome sequence
RT   of uropathogenic Escherichia coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX   MEDLINE=93023838; PubMed=1406252;
RA   Hromockyj A.E., Tucker S.C., Maurelli A.T.;
RT   "Temperature regulation of Shigella virulence: identification of the
RT   repressor gene virR, an analogue of hns, and partial complementation
```

```
RT   by tyrosyl transfer RNA (tRNA1(Tyr)).";
RL   Mol. Microbiol. 6:2113-2124(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX   MEDLINE=22272406; PubMed=12384590;
RA   Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA   Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA   Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA   Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA   Yu J.;
RT   "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT   through comparison with genomes of Escherichia coli K12 and O157.";
RL   Nucleic Acids Res. 30:4432-4441(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX   MEDLINE=22590274; PubMed=12704152;
RA   Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA   Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA   Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA   Schwartz D.C., Blattner F.R.;
RT   "Complete genome sequence and comparative genomics of Shigella
RT   flexneri serotype 2a strain 2457T.";
RL   Infect. Immun. 71:2775-2786(2003).
RN   [5]
RP   IDENTIFICATION.
RX   MEDLINE=94042872; PubMed=8226647;
RA   Nagy P.L., McCorkle G., Zalkin H.;
RT   "purU, a source of formate for purT-dependent phosphoribosyl-N-
RT   formylglycinamide synthesis.";
RL   J. Bacteriol. 175:7066-7073(1993).
CC   -!- FUNCTION: PRODUCES FORMATE FROM FORMYL-TETRAHYDROFOLATE. PROVIDES
CC       THE MAJOR SOURCE OF FORMATE FOR THE PURT-DEPENDENT SYNTHESIS OF
CC       5'-PHOSPHORIBOSYL-N-FORMYLGLYCINAMIDE (FGAR) DURING AEROBIC
CC       GROWTH. HAS A ROLE IN REGULATING THE ONE-CARBON POOL.
CC   -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + H(2)O = formate +
CC       tetrahydrofolate.
CC   -!- ENZYME REGULATION: Activated by methionine, inhibited by glycine
CC       (By similarity).
CC   -!- PATHWAY: De novo purine biosynthesis.
CC   -!- SUBUNIT: Homohexamer (By similarity).
CC   -!- SIMILARITY: SOME, TO GAR TRANSFORMYLASE (PURN).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE016760; AAN80163.1; -.
DR   EMBL; X66849; -; NOT_ANNOTATED_CDS.
DR   EMBL; AE015150; AAN42845.1; -.
DR   EMBL; AE016982; AAP16730.1; -.
DR   HSSP; P08179; 1GRC.
DR   InterPro; IPR002912; ACT.
DR   InterPro; IPR002376; formyl_transf.
DR   InterPro; IPR004810; PurU.
DR   Pfam; PF01842; ACT; 1.
DR   Pfam; PF00551; formyl_transf; 1.
DR   PRINTS; PR01575; FFH4HYDRLASE.
DR   TIGRFAMs; TIGR00655; PurU; 1.
KW   Purine biosynthesis; Hydrolase; One-carbon metabolism;
KW   Complete proteome.
FT   ACT_SITE    225    225       BY SIMILARITY.
FT   CONFLICT     44     44       R -> L (IN REF. 2).
SQ   SEQUENCE   280 AA;  31920 MW;  55BC16B62727A419 CRC64;

  Query Match             44.3%;  Score 47;  DB 1;  Length 280;
  Best Local Similarity   52.9%;  Pred. No. 7.4;
  Matches    9;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :|  |  | | ||
Db        193 LPAFIGARPYHQAYERG 209


RESULT 4
PURU_ECOLI
ID   PURU_ECOLI     STANDARD;      PRT;   280 AA.
AC   P37051;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-FH(4)
DE   hydrolase).
GN   PURU OR TGS OR B1232.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-6.
RC   STRAIN=K12;
RX   MEDLINE=94042872; PubMed=8226647;
RA   Nagy P.L., McCorkle G., Zalkin H.;
RT   "purU, a source of formate for purT-dependent phosphoribosyl-N-
RT   formylglycinamide synthesis.";
RL   J. Bacteriol. 175:7066-7073(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=94110230; PubMed=8282700;
RA   Boesl M., Kersten H.;
RT   "Organization and functions of genes in the upstream region of tyrT
RT   of Escherichia coli: phenotypes of mutants with partial deletion of a
RT   new gene (tgs).";
RL   J. Bacteriol. 176:221-231(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [5]
RP   CHARACTERIZATION.
RX   MEDLINE=95173107; PubMed=7868604;
RA   Nagy P.L., Marolewski A., Benkovic S.J., Zalkin H.;
RT   "Formyltetrahydrofolate hydrolase, a regulatory enzyme that functions
RT   to balance pools of tetrahydrofolate and one-carbon tetrahydrofolate
RT   adducts in Escherichia coli.";
RL   J. Bacteriol. 177:1292-1298(1995).
CC   -!- FUNCTION: PRODUCES FORMATE FROM FORMYL-TETRAHYDROFOLATE. PROVIDES
CC       THE MAJOR SOURCE OF FORMATE FOR THE PURT-DEPENDENT SYNTHESIS OF
CC       5'-PHOSPHORIBOSYL-N-FORMYLGLYCINAMIDE (FGAR) DURING AEROBIC
CC       GROWTH. HAS A ROLE IN REGULATING THE ONE-CARBON POOL.
CC   -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + H(2)O = formate +
```

```
CC       tetrahydrofolate.
CC   -!- ENZYME REGULATION: Activated by methionine, inhibited by glycine.
CC   -!- PATHWAY: De novo purine biosynthesis.
CC   -!- SUBUNIT: Homohexamer.
CC   -!- SIMILARITY: SOME, TO GAR TRANSFORMYLASE (PURN).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L20251; AAC36846.1; -.
DR   EMBL; M64675; AAA16860.1; ALT_INIT.
DR   EMBL; AE000221; AAC74314.1; -.
DR   EMBL; D90758; BAA36100.1; -.
DR   EMBL; D90759; BAA36112.1; -.
DR   EMBL; D90852; BAA16026.1; -.
DR   PIR; C36871; C36871.
DR   EcoGene; EG11819; purU.
DR   InterPro; IPR002912; ACT.
DR   InterPro; IPR002376; formyl_transf.
DR   InterPro; IPR004810; PurU.
DR   Pfam; PF01842; ACT; 1.
DR   Pfam; PF00551; formyl_transf; 1.
DR   PRINTS; PR01575; FFH4HYDRLASE.
DR   TIGRFAMs; TIGR00655; PurU; 1.
KW   Purine biosynthesis; Hydrolase; One-carbon metabolism;
KW   Complete proteome.
FT   ACT_SITE    225    225       BY SIMILARITY.
SQ   SEQUENCE   280 AA;  31934 MW;  5667406D2727A2C2 CRC64;

  Query Match             44.3%;  Score 47;  DB 1;  Length 280;
  Best Local Similarity   52.9%;  Pred. No. 7.4;
  Matches    9;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :|  |  | | ||
Db        193 LPAFIGARPYHQAYERG 209


RESULT 5
COBI_MYCTU
ID   COBI_MYCTU     STANDARD;      PRT;   508 AA.
AC   Q10677;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Cobalamin biosynthesis protein COBIJ [Includes: Precorrin-2 C20-
DE   methyltransferase (EC 2.1.1.130) (S-adenosyl-L-methionine--precorrin-2
DE   methyltransferase) (SP2MT); Precorrin-3 methylase (EC 2.1.1.-)].
GN   COBIJ OR COBI OR RV2066 OR MT2126 OR MTCY49.05 OR MB2092.
OS   Mycobacterium tuberculosis, and
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773, 1765;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX   MEDLINE=22206494; PubMed=12218036;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT   "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   J. Bacteriol. 184:5479-5490(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.bovis; STRAIN=AF2122/97;
RX   MEDLINE=22709107; PubMed=12788972;
RA   Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA   Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA   Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA   Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT   "The complete genome sequence of Mycobacterium bovis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC   -!- FUNCTION: METHYLATES PRECORRIN-2 AT THE C-20 POSITION TO PRODUCE
CC       PRECORRIN-3A (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-2 = S-
CC       adenosyl-L-homocysteine + precorrin-3A.
CC   -!- PATHWAY: Cobalamin biosynthesis.
CC   -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS SUMT, CYSG, CBIF/COBM
CC       AND CBIL/COBI.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z73966; CAA98214.1; -.
DR   EMBL; AE007063; AAK46406.1; -.
DR   EMBL; BX248341; CAD96945.1; -.
DR   PIR; E70764; E70764.
DR   TIGR; MT2126; -.
DR   TubercuList; Rv2066; -.
DR   InterPro; IPR006364; CobI_CbiL.
DR   InterPro; IPR006363; CobJ.
DR   InterPro; IPR000878; Cor/por_Metransf.
DR   InterPro; IPR003043; Uropor_Metransf.
DR   Pfam; PF00590; TP_methylase; 2.
DR   TIGRFAMs; TIGR01467; cobI_cbiL; 1.
DR   TIGRFAMs; TIGR01466; cobJ_cbiH; 1.
DR   PROSITE; PS00839; SUMT_1; 1.
DR   PROSITE; PS00840; SUMT_2; 1.
KW   Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW   Methyltransferase; Multifunctional enzyme; Complete proteome.
FT   DOMAIN       1    243       PRECORRIN-2 C20-METHYLTRANSFERASE.
FT   DOMAIN     244    508       PRECORRIN-3 METHYLASE.
SQ   SEQUENCE   508 AA;  53910 MW;  95AC066F022C4DC1 CRC64;

  Query Match             43.4%;  Score 46;  DB 1;  Length 508;
  Best Local Similarity   47.1%;  Pred. No. 19;
  Matches    8;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 MAGLPAVVGLSPGEQEY 17
              : |  ||||| ||: ::
Db        245 LTGTVAVVGLGPGDSDW 261


RESULT 6
CNA2_MOUSE
ID   CNA2_MOUSE     STANDARD;      PRT;  1172 AA.
AC   Q80XC6; Q8R3D7; Q99LT9;
```

```
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Protein C14orf102 homolog.
GN   C14ORF102.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   CONCEPTUAL TRANSLATION, AND RECONSTRUCTION FROM EST.
RA   Reynaud S.;
RL   Unpublished observations (AUG-2003).
RN   [2]
RP   SEQUENCE OF 21-1172 FROM N.A.
RC   TISSUE=Breast tumor, and Olfactory epithelium;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SIMILARITY: Contains 5 HAT repeats.
CC   -!- CAUTION: This is a conceptual translation; a frameshift was
CC       introduced in position 28 to produce the correct N-terminus, and
CC       to extend the similarity with the human ortholog.
CC   -!- CAUTION: Ref.1 (AAH51175) sequence differs from that shown due to
CC       a stop codon in position 1146 which was translated as Gln to
CC       extend the sequence.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; BY182441; -; NOT_ANNOTATED_CDS.
DR   EMBL; BC002230; AAH02230.1; -.
DR   EMBL; BC025577; AAH25577.1; ALT_INIT.
DR   EMBL; BC051175; AAH51175.1; ALT_SEQ.
DR   InterPro; IPR003107; HAT.
DR   SMART; SM00386; HAT; 5.
KW   Repeat; Coiled coil.
FT   DOMAIN       78    330       COILED COIL (POTENTIAL).
FT   REPEAT      314    346       HAT 1.
FT   REPEAT      404    436       HAT 2.
FT   REPEAT      766    800       HAT 3.
FT   REPEAT      986   1018       HAT 4.
FT   REPEAT     1075   1109       HAT 5.
FT   DOMAIN       86     93       POLY-LYS.
FT   DOMAIN      146    154       POLY-ALA.
FT   CONFLICT    707    707       R -> H (IN REF. 1; AAH25577).
FT   CONFLICT    741    742       LS -> VC (IN REF. 1; AAH25577).
SQ   SEQUENCE   1172 AA;  133424 MW;  ABE37A8BCE7A528F CRC64;

  Query Match           41.5%;  Score 44;  DB 1;  Length 1172;
  Best Local Similarity   53.3%;  Pred. No. 94;
  Matches    8;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          5 PAVVGLSPGEQEYHR 19
              | :  |  |||| ||
Db        351 PGIYALGEGEQEKHR 365


RESULT 7
MINT_HUMAN
ID   MINT_HUMAN     STANDARD;      PRT;  3664 AA.
AC   Q96T58; Q9H9A8; Q9NWH5; Q9UQ01; Q9Y556;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Msx2-interacting protein (SMART/HDAC1 associated repressor protein).
GN   MINT OR SHARP OR KIAA0929.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, INDUCTION,
RP   RNA-BINDING, AND INTERACTION WITH NCOR2; HDAC1; HDAC2; RBBP4; MBD3;
RP   RAR AND MTA1L1.
RC   TISSUE=Liver, and Pituitary;
RX   MEDLINE=21231190; PubMed=11331609;
RA   Shi Y., Downes M., Xie W., Kao H.-Y., Ordentlich P., Tsai C.-C.,
RA   Hon M., Evans R.M.;
RT   "Sharp, an inducible cofactor that integrates nuclear receptor
RT   repression and activation.";
RL   Genes Dev. 15:1140-1151(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Bird C.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 294-3664 FROM N.A.
RA   Rhodes S., Huckle E.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 793-1595 FROM N.A., AND VARIANT PRO-1091.
RC   TISSUE=Embryo, and Teratocarcinoma;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA   Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA   Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA   Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 2002-3664 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99246063; PubMed=10231032;
RA   Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosawa M.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIII.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:63-70(1999).
RN   [6]
RP   INTERACTION WITH PPARD.
RX   MEDLINE=21874127; PubMed=11867749;
RA   Shi Y., Hon M., Evans R.M.;
RT   "The peroxisome proliferator-activated receptor delta, an integrator
RT   of transcriptional repression and nuclear receptor signaling.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:2613-2618(2002).
RN   [7]
RP   FUNCTION, AND INTERACTION WITH RBPSUH.
RX   MEDLINE=22261914; PubMed=12374742;
RA   Oswald F., Kostezka U., Astrahantseff K., Bourteele S., Dillinger K.,
RA   Zechner U., Ludwig L., Wilda M., Hameister H., Knoechel W., Liptay S.,
RA   Schmid R.M.;
```

```
RT   "SHARP is a novel component of the Notch/RBP-Jkappa signalling
RT   pathway.";
RL   EMBO J. 21:5417-5426(2002).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF SPOC DOMAIN.
RX   MEDLINE=22777836; PubMed=12897056;
RA   Ariyoshi M., Schwabe J.W.R.;
RT   "A conserved structural motif reveals the essential transcriptional
RT   repression function of Spen proteins and their role in developmental
RT   signaling.";
RL   Genes Dev. 17:1909-1920(2003).
CC   -!- FUNCTION: Essential corepressor protein, which probably regulates
CC       different key pathways such as the Notch pathway. Negative
CC       regulator of the Notch pathway via its interaction with RBPSUH,
CC       which prevents the association between NOTCH1 and RBPSUH, and
CC       therefore suppresses the transactivation activity of Notch
CC       signaling. Blocks the differentiation of precursor B cells into
CC       marginal zone B cells. Probably represses transcription via the
CC       recruitment of large complexes containing histone deacetylase
CC       proteins. May bind both to DNA and RNA.
CC   -!- SUBUNIT: Interacts with MSX2 (By similarity). Interacts with
CC       NCOR2, HDAC1, HDAC2, RBBP4, MBD3 and MTA1L1. Interacts with
CC       RBPSUH; this interaction may prevent the interaction between
CC       RBPSUH and NOTCH1. Interacts with the nuclear receptors RAR and
CC       PPARD. Interacts with RAR in absence of ligand. Bind to the
CC       steroid receptor RNA coactivator SRA.
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associates with chromatin.
CC   -!- TISSUE SPECIFICITY: Expressed at high level in brain, testis,
CC       spleen and thymus. Expressed at intermediate level in kidney,
CC       liver, mammary gland and skin.
CC   -!- INDUCTION: By hormone 17-beta-estradiol (E2).
CC   -!- DOMAIN: The RID domain mediates the interaction with nuclear
CC       receptors (By similarity).
CC   -!- DOMAIN: The SPOC domain, which mediates the interaction with
CC       NCOR2, is essential for the repressive activity.
CC   -!- SIMILARITY: Belongs to the Spen family.
CC   -!- SIMILARITY: Contains 1 RID (receptor interacting) domain.
CC   -!- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
CC   -!- SIMILARITY: Contains 1 SPOC domain.
CC   -!- CAUTION: Ref.2 sequences differ from that shown due to erroneous
CC       gene model prediction.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF356524; AAK52750.1; -.
DR   EMBL; AL034555; CAB85442.1; ALT_SEQ.
DR   EMBL; AL034555; CAB85444.1; ALT_SEQ.
DR   EMBL; AL450998; -; NOT_ANNOTATED_CDS.
DR   EMBL; AL096858; CAB51072.1; ALT_INIT.
DR   EMBL; AK000882; BAA91405.1; ALT_INIT.
DR   EMBL; AK022949; BAB14324.1; ALT_INIT.
DR   EMBL; AB023146; BAA76773.1; -.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   PDB; 1OW1; 19-AUG-03.
DR   Pfam; PF00076; rrm; 4.
DR   SMART; SM00360; RRM; 4.
DR   PROSITE; PS50102; RRM; 4.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
DR   PROSITE; PS50917; SPOC; 1.
KW   Transcription regulation; Repressor; Nuclear protein; DNA-binding;
KW   Repeat; RNA-binding; Coiled coil; 3D-structure; Polymorphism.
FT   DOMAIN        1    573       DNA-BINDING (BY SIMILARITY).
FT   DOMAIN        6     81       RNA-BINDING (RRM) 1.
FT   DOMAIN      335    415       RNA-BINDING (RRM) 2.
FT   DOMAIN      438    513       RNA-BINDING (RRM) 3.
FT   DOMAIN      517    589       RNA-BINDING (RRM) 4.
FT   DOMAIN      688    715       COILED COIL (POTENTIAL).
FT   DOMAIN      977   1004       COILED COIL (POTENTIAL).
FT   DOMAIN     1170   1191       COILED COIL (POTENTIAL).
FT   DOMAIN     1408   1428       COILED COIL (POTENTIAL).
FT   DOMAIN     1496   1529       COILED COIL (POTENTIAL).
FT   DOMAIN     1592   1612       COILED COIL (POTENTIAL).
FT   DOMAIN     1928   1944       COILED COIL (POTENTIAL).
FT   DOMAIN     2201   2707       RID.
FT   DOMAIN     3498   3664       SPOC.
FT   DOMAIN     2130   2464       INTERACTION WITH MSX2 (BY SIMILARITY).
FT   DOMAIN     2709   2870       INTERACTION WITH RBPSUH (BY SIMILARITY).
FT   DOMAIN      125    277       ARG-RICH.
FT   DOMAIN      240    325       SER-RICH.
FT   DOMAIN      616    810       ARG-RICH.
FT   DOMAIN      624    697       TYR-RICH.
FT   DOMAIN     2428   2520       PRO-RICH.
FT   DOMAIN     3220   3482       PRO-RICH.
FT   VARIANT     970    970       A -> V (in dbSNP:848208).
FT                                /FTId=VAR_017119.
FT   VARIANT    1091   1091       L -> P (in dbSNP:848209).
FT                                /FTId=VAR_017120.
FT   VARIANT    2360   2360       N -> D (in dbSNP:848210).
FT                                /FTId=VAR_017121.
FT   CONFLICT    956    956       G -> D (IN REF. 4).
SQ   SEQUENCE   3664 AA;  402245 MW;  5228C58533E5B27B CRC64;

  Query Match             41.5%;  Score 44;  DB 1;  Length 3664;
  Best Local Similarity   47.1%;  Pred. No. 3.1e+02;
  Matches    8;  Conservative    3;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          2 AGLPAVVGLSPGEQEYH 18
              ||:||:    | |:| |
Db       3145 AGVPALASQHPPEEEVH 3161


RESULT 8
COBL_METJA
ID   COBL_METJA     STANDARD;      PRT;   211 AA.
AC   Q58917;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Probable precorrin-6Y C5,15-methyltransferase [decarboxylating]
DE   (EC 2.1.1.132) (Precorrin-6 methyltransferase) (Precorrin-6Y
DE   methylase).
GN   COBL OR MJ1522.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- FUNCTION: CATALYZES THE METHYLATION OF BOTH C-5 AND C-15 IN
CC       PRECORRIN-6Y TO FORM PRECORRIN-8X (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 2 S-adenosyl-L-methionine + precorrin-6Y = 2
CC       S-adenosyl-L-homocysteine + precorrin-8X + CO(2).
CC   -!- PATHWAY: Cobalamin biosynthesis.
CC   -!- SIMILARITY: TO S.TYPHIMURIUM CBIE; ALSO, LOW, TO OTHER
CC       METHYLASES INVOLVED IN COBALAMIN BIOSYNTHESIS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U67593; AAB99541.1; -.
DR   PIR; A64490; A64490.
DR   TIGR; MJ1522; -.
DR   InterPro; IPR000878; Cor/por_Metransf.
DR   Pfam; PF00590; TP_methylase; 1.
KW   Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW   Methyltransferase; Complete proteome.
SQ   SEQUENCE   211 AA;  23805 MW;  279A1A2B14369510 CRC64;

  Query Match             40.6%;  Score 43;  DB 1;  Length 211;
  Best Local Similarity   54.5%;  Pred. No. 23;
  Matches    6;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          7 VVGLSPGEQEY 17
              :||: ||::||
Db          4 IVGIGPGDREY 14


RESULT 9
MER_METTI
ID   MER_METTI      STANDARD;      PRT;   326 AA.
AC   Q9UXP0;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Coenzyme F420-dependent N(5),N(10)-methylenetetrahydromethanopterin
DE   reductase (EC 1.5.99.11) (Methylene-H(4)MPT reductase).
GN   MER OR FFDA.
OS   Methanolobus tindarius.
OC   Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC   Methanosarcinaceae; Methanolobus.
OX   NCBI_TaxID=2221;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 2278;
RX   MEDLINE=99132696; PubMed=9933933;
RA   Westenberg D.J., Braune A., Ruppert C., Mueller V., Herzberg C.,
RA   Gottschalk G., Blaut M.;
RT   "The F420H2-dehydrogenase from Methanolobus tindarius: cloning of the
RT   ffd operon and expression of the genes in Escherichia coli.";
RL   FEMS Microbiol. Lett. 170:389-398(1999).
CC   -!- FUNCTION: Catalyzes the reversible reduction of methylene-H(4)MPT
CC       to methyl-H(4)MPT (By similarity).
CC   -!- CATALYTIC ACTIVITY: N(5),N(10)-methylenetetrahydromethanopterin +
CC       reduced coenzyme F420 = 5-methyl-5,6,7,8-tetrahydromethanopterin +
CC       coenzyme F420.
CC   -!- PATHWAY: Methanogenesis from carbon dioxide; fifth step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the mer family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AJ011519; CAB56639.1; -.
DR   PIR; T45226; T45226.
DR   HAMAP; MF_01091; -; 1.
DR   InterPro; IPR002103; Bac_luciferase.
DR   Pfam; PF00296; bac_luciferase; 1.
KW   Methanogenesis; One-carbon metabolism; Oxidoreductase.
SQ   SEQUENCE   326 AA;  34043 MW;  16F3AB9733A45D82 CRC64;

  Query Match             40.6%;  Score 43;  DB 1;  Length 326;
  Best Local Similarity   53.3%;  Pred. No. 35;
  Matches    8;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MAGLPAVVGLSPGEQ 15
              ::|  |::|| ||||
Db         83 ISGGRAILGLGPGEQ 97


RESULT 10
COBL_MYCTU
ID   COBL_MYCTU     STANDARD;      PRT;   390 AA.
AC   Q10671;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Precorrin-6Y C5,15-methyltransferase [decarboxylating] (EC 2.1.1.132)
DE   (Precorrin-6 methyltransferase) (Precorrin-6Y methylase).
GN   COBL OR RV2072C OR MT2132 OR MTCY49.11C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RX   MEDLINE=22206494; PubMed=12218036;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT   "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   J. Bacteriol. 184:5479-5490(2002).
CC   -!- FUNCTION: CATALYZES THE METHYLATION OF BOTH C-5 AND C-15 IN
CC       PRECORRIN-6Y TO FORM PRECORRIN-8X.
CC   -!- CATALYTIC ACTIVITY: 2 S-adenosyl-L-methionine + precorrin-6Y = 2
CC       S-adenosyl-L-homocysteine + precorrin-8X + CO(2).
CC   -!- PATHWAY: Cobalamin biosynthesis.
CC   -!- SIMILARITY: TO S.TYPHIMURIUM CBIE; ALSO, LOW, TO OTHER
CC       METHYLASES INVOLVED IN COBALAMIN BIOSYNTHESIS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z73966; CAA98225.1; -.
DR   EMBL; AE007063; AAK46412.1; -.
DR   PIR; C70765; C70765.
DR   TIGR; MT2132; -.
DR   TubercuList; Rv2072c; -.
DR   InterPro; IPR006365; CobL.
DR   InterPro; IPR000878; Cor/por_Metransf.
```

```
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF00590; TP_methylase; 1.
DR   TIGRFAMs; TIGR01468; cobL_cbiET; 1.
KW   Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW   Methyltransferase; Complete proteome.
FT   CONFLICT    205    205       L -> P (IN REF. 2).
FT   CONFLICT    327    327       D -> H (IN REF. 2).
SQ   SEQUENCE   390 AA;  41854 MW;  FB42EFF7562F21F3 CRC64;

  Query Match             40.6%;  Score 43;  DB 1;  Length 390;
  Best Local Similarity   60.0%;  Pred. No. 43;
  Matches    9;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYH 18
              |||| ||||   : |
Db         55 LPAVQGLSPDGADLH 69


RESULT 11
YT6J_CAEEL
ID   YT6J_CAEEL     STANDARD;      PRT;   505 AA.
AC   P90838;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical UPF0027 protein F16A11.2 in chromosome I.
GN   F16A11.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Baynes C.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   REVISIONS.
RA   Durbin R.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the UPF0027 (rtcB) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z81505; CAB04121.2; -.
DR   WormPep; F16A11.2; CE23663.
DR   InterPro; IPR001233; UPF0027.
DR   Pfam; PF01139; UPF0027; 1.
DR   PROSITE; PS01288; UPF0027; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   505 AA;  55230 MW;  D528F702E2586909 CRC64;

  Query Match             40.6%;  Score 43;  DB 1;  Length 505;
  Best Local Similarity   45.0%;  Pred. No. 56;
  Matches    9;  Conservative    3;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 MAGLPAVVGLSPGEQEYHRG 20
              :| || :|| | |  : | |
Db         72 VASLPGIVGHSIGLPDIHSG 91


RESULT 12
C4AD_DROME
ID   C4AD_DROME     STANDARD;      PRT;   516 AA.
AC   Q9V4T3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable cytochrome P450 4ad1 (EC 1.14.-.-) (CYPIVAD1).
GN   CYP4AD1 OR CG2110.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley; TISSUE=Embryo;
RX   MEDLINE=22426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC   -!- FUNCTION: May be involved in the metabolism of insect hormones and
CC       in the breakdown of synthetic insecticides (By similarity).
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum
CC       (Potential).
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; AE003837; AAF59092.1; -.
DR   EMBL; AY061058; AAL28606.1; -.
DR   HSSP; P14779; 1JPZ.
DR   FlyBase; FBgn0033292; Cyp4ad1.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW   Endoplasmic reticulum; Hypothetical protein.
FT   METAL       445    445       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   516 AA;  58870 MW;  648EA22492AF58C7 CRC64;

  Query Match             40.6%;  Score 43;  DB 1;  Length 516;
  Best Local Similarity   56.2%;  Pred. No. 57;
  Matches    9;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHR 19
              |||| || ||  :: |
Db        469 LPAVDGLPPGINDHSR 484


RESULT 13
CH60_PROAC
ID   CH60_PROAC     STANDARD;      PRT;   544 AA.
AC   Q9K2U4;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein) (Heat shock protein
DE   60).
GN   GROL OR GROEL OR HSP60.
OS   Propionibacterium acnes.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Propionibacterineae; Propionibacteriaceae; Propionibacterium.
OX   NCBI_TaxID=1747;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cho Y.;
RT   "Propionibacterium acnes hsp60 gene.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=P37;
RA   Farrar M.D., Ingham E., Holland K.T.;
RT   "Cloning and sequencing of a groEL homolog from Propionibacterium
RT   acnes.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; AB036414; BAA92382.1; -.
DR   EMBL; AF222061; AAF33788.1; -.
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   InterPro; IPR008950; GroEL-ATPase.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding.
SQ   SEQUENCE   544 AA;  56839 MW;  9B4314A8BF8A94DC CRC64;

  Query Match             40.6%;  Score 43;  DB 1;  Length 544;
  Best Local Similarity   52.9%;  Pred. No. 60;
  Matches    9;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 MAGLPAVVGLSPGEQEY 17
              :|||||  ||:    ||
Db        466 VAGLPAGQGLNAANDEY 482


RESULT 14
HNFA_MOUSE
ID   HNFA_MOUSE     STANDARD;      PRT;   628 AA.
AC   P22361;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hepatocyte nuclear factor 1-alpha (HNF-1A) (Liver-specific
DE   transcription factor LF-B1) (LFB1).
GN   TCF1 OR HNF1A OR HNF-1A OR HNF-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91088607; PubMed=2263635;
RA   Kuo C.J., Conley P.B., Hsieh C.L., Francke U., Crabtree G.R.;
RT   "Molecular cloning, functional expression, and chromosomal
RT   localization of mouse hepatocyte nuclear factor 1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:9838-9842(1990).
CC   -!- FUNCTION: Required for the expression of several liver specific
CC       genes. Binds to the inverted palindrome 5'-GTTAATNATTAAC-3'.
CC   -!- SUBUNIT: Binds DNA as a dimer.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Liver.
CC   -!- SIMILARITY: Belongs to the HNF1 homeobox family.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; M57966; AAA37821.1; -.
DR   PIR; A39262; A39262.
DR   PDB; 1F93; 20-SEP-00.
DR   PDB; 1G2Y; 17-JAN-01.
DR   PDB; 1G2Z; 17-JAN-01.
DR   PDB; 1G39; 17-JAN-01.
DR   PDB; 1JB6; 11-JUL-01.
DR   PDB; 1LFB; 31-OCT-93.
DR   TRANSFAC; T01211; -.
DR   MGD; MGI:98504; Tcf1.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   InterPro; IPR006899; HNF-1_N.
DR   InterPro; IPR006898; HNF1a_C.
DR   InterPro; IPR006897; HNF1b_C.
DR   InterPro; IPR001356; Homeobox.
DR   Pfam; PF04814; HNF-1_N; 1.
DR   Pfam; PF04813; HNF-1A_C; 1.
DR   Pfam; PF04812; HNF-1B_C; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
```

```
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Transcription regulation; DNA-binding; Homeobox; Nuclear protein;
KW   Activator; Trans-acting factor; 3D-structure.
FT   DOMAIN        1     31       DIMERIZATION.
FT   DOMAIN       71     80       ASP/GLU-RICH (ACIDIC) (POTENTIAL
FT                                INVOLVEMENT WITH TRANSCRIPTION).
FT   DOMAIN      197    205       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DNA_BIND    199    279       HOMEOBOX.
FT   DOMAIN      238    258       21 AMINO ACID LOOP BETWEEN HELIX 2 AND 3.
SQ   SEQUENCE   628 AA;  67237 MW;  737920D1A369B9DD CRC64;

  Query Match             40.6%;  Score 43;  DB 1;  Length 628;
  Best Local Similarity   57.1%;  Pred. No. 70;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 MAGLPAVVGLSPGE 14
              || || |: : |||
Db        404 MASLPGVMTIGPGE 417


RESULT 15
HNFA_RAT
ID   HNFA_RAT       STANDARD;      PRT;   628 AA.
AC   P15257;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hepatocyte nuclear factor 1-alpha (HNF-1A) (Liver-specific
DE   transcription factor LF-B1) (LFB1).
GN   TCF1 OR HNF1A OR HNF-1A OR HNF-1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=90003224; PubMed=2571419;
RA   Frain M., Swart G., Monaci P., Nicosia A., Staempfli S., Frank R.,
RA   Cortese R.;
RT   "The liver-specific transcription factor LF-B1 contains a highly
RT   diverged homeobox DNA binding domain.";
RL   Cell 59:145-157(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91016926; PubMed=2216777;
RA   Chouard T., Blumenfeld M., Bach I., Vandekerckhove J., Cereghini S.,
RA   Yaniv M.;
RT   "A distal dimerization domain is essential for DNA-binding by the
RT   atypical HNF1 homeodomain.";
RL   Nucleic Acids Res. 18:5853-5863(1990).
RN   [3]
RP   SEQUENCE OF 166-628 FROM N.A.
RX   MEDLINE=90249741; PubMed=1970973;
RA   Baumhueter S., Mendel D.B., Conley P.B., Kuo C.J., Turk C.,
RA   Graves M.K., Edwards C.A., Courtois G., Crabtree G.R.;
RT   "HNF-1 shares three sequence motifs with the POU domain proteins and
RT   is identical to LF-B1 and APF.";
RL   Genes Dev. 4:372-379(1990).
RN   [4]
RP   SEQUENCE OF 1-12 FROM N.A.
RC   STRAIN=Sprague-Dawley;
RA   Tomei L., Piaggio G., Toniatti C., Lazzaro D., de Francesco R.,
RA   Pozzi L., Gerstner J., Cortese R.;
RL   Submitted (AUG-1992) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   POSITION OF HOMEOBOX.
RX   MEDLINE=90106643; PubMed=1967225;
RA   Finney M.;
RT   "The homeodomain of the transcription factor LF-B1 has a 21 amino
RT   acid loop between helix 2 and helix 3.";
RL   Cell 60:5-6(1990).
RN   [6]
RP   STRUCTURE BY NMR OF 1-32.
RX   MEDLINE=91105074; PubMed=1988016;
RA   Pastore A., de Francesco R., Barbato G., Castiglione Morelli M.A.,
RA   Motta A., Cortese R.;
RT   "1H resonance assignment and secondary structure determination of the
RT   dimerization domain of transcription factor LFB1.";
RL   Biochemistry 30:148-153(1991).
RN   [7]
RP   STRUCTURE BY NMR OF 195-286.
RX   MEDLINE=93259120; PubMed=8491172;
RA   Leiting B., de Francesco R., Tomei L., Cortese R., Otting G.,
RA   Wuethrich K.;
RT   "The three-dimensional NMR-solution structure of the polypeptide
RT   fragment 195-286 of the LFB1/HNF1 transcription factor from rat liver
RT   comprises a nonclassical homeodomain.";
RL   EMBO J. 12:1797-1803(1993).
RN   [8]
RP   STRUCTURE BY NMR OF 195-286.
RX   MEDLINE=97272000; PubMed=9126845;
RA   Schott O., Billeter M., Leiting B., Wider G., Wuethrich K.;
RT   "The NMR solution structure of the non-classical homeodomain from the
RT   rat liver LFB1/HNF1 transcription factor.";
RL   J. Mol. Biol. 267:673-683(1997).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 195-286.
RX   MEDLINE=93259121; PubMed=8491173;
RA   Ceska T.A., Lamers M., Monaci P., Nicosia A., Cortese R., Suck D.;
RT   "The X-ray structure of an atypical homeodomain present in the rat
RT   liver transcription factor LFB1/HNF1 and implications for DNA
RT   binding.";
RL   EMBO J. 12:1805-1810(1993).
CC   -!- FUNCTION: Required for the expression of several liver specific
CC       genes. Binds to the inverted palindrome 5'-GTTAATNATTAAC-3'.
CC   -!- SUBUNIT: Binds DNA as a dimer.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Liver.
CC   -!- SIMILARITY: Belongs to the HNF1 homeobox family.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J03170; AAA41524.1; -.
DR   EMBL; X54423; CAA38295.1; -.
DR   EMBL; X67649; CAA47891.1; -.
DR   EMBL; X53297; CAA37387.1; ALT_INIT.
DR   PIR; A33333; A33333.
DR   PIR; S25485; S25485.
DR   PDB; 1LFB; 31-OCT-93.
DR   PDB; 2LFB; 12-MAR-97.
DR   TRANSFAC; T00369; -.
DR   InterPro; IPR006899; HNF-1_N.
DR   InterPro; IPR006898; HNF1a_C.
DR   InterPro; IPR006897; HNF1b_C.
DR   InterPro; IPR001356; Homeobox.
DR   Pfam; PF04814; HNF-1_N; 1.
DR   Pfam; PF04813; HNF-1A_C; 1.
DR   Pfam; PF04812; HNF-1B_C; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Transcription regulation; DNA-binding; Homeobox; Nuclear protein;
KW   Activator; Trans-acting factor; 3D-structure.
FT   DOMAIN        1     31       DIMERIZATION.
FT   DOMAIN       71     80       ASP/GLU-RICH (ACIDIC) (POTENTIAL
FT                                INVOLVEMENT WITH TRANSCRIPTION).
FT   DOMAIN      197    205       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DNA_BIND    199    279       HOMEOBOX.
```

```
FT   DOMAIN       238    258       21 AMINO ACID LOOP BETWEEN HELIX 2 AND 3.
FT   HELIX        208    218
FT   TURN         219    220
FT   HELIX        226    241
FT   TURN         242    245
FT   TURN         248    249
FT   TURN         251    254
FT   HELIX        255    257
FT   HELIX        261    273
FT   TURN         274    275
SQ   SEQUENCE   628 AA;  67213 MW;  8D28099308C86A52 CRC64;

  Query Match             40.6%;  Score 43;  DB 1;  Length 628;
  Best Local Similarity   57.1%;  Pred. No. 70;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 MAGLPAVVGLSPGE 14
              || || |: : |||
Db        404 MASLPGVMTIGPGE 417


Search completed: May  7, 2004, 12:34:31
Job time : 8.2 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:26:40 ; Search time 37.4 Seconds
(without alignments)
168.726 Million cell updates/sec

Title: US-09-786-214A-9
Perfect score: 106
Sequence: 1 MAGLPAVVGLSPGEQEYHRG 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	64	60.4	41	11	Q8K408	Q8k408 rattus norv
2	52	49.1	150	16	Q8PE55	Q8pe55 xanthomonas
3	50	47.2	242	16	Q89Y25	Q89y25 bradyrhizob
4	50	47.2	277	16	Q9KQK6	Q9kqk6 vibrio chol
5	50	47.2	284	16	Q7VFI5	Q7vfi5 helicobacte
6	50	47.2	381	16	Q9RRL5	Q9rrl5 deinococcus
7	49	46.2	277	16	Q87RD3	Q87rd3 vibrio para
8	49	46.2	4823	13	O93321	O93321 fugu rubrip
9	48	45.3	192	15	Q8UTQ8	Q8utq8 human immun
10	48	45.3	192	15	Q90D29	Q90d29 human immun
11	48	45.3	278	16	Q9CMF6	Q9cmf6 pasteurella
12	48	45.3	300	16	Q8EMZ7	Q8emz7 oceanobacil
13	48	45.3	485	10	Q7XNE4	Q7xne4 oryza sativ
14	47	44.3	192	15	Q994Q2	Q994q2 human immun
15	47	44.3	192	15	Q9QSR2	Q9qsr2 human immun
16	47	44.3	192	15	Q994B2	Q994b2 human immun
17	47	44.3	250	16	Q8D0G6	Q8d0g6 yersinia pe
18	47	44.3	280	16	Q8ZP42	Q8zp42 salmonella
19	47	44.3	280	16	Q8XDC4	Q8xdc4 escherichia
20	47	44.3	280	16	Q8Z7F3	Q8z7f3 salmonella
21	47	44.3	289	16	Q8KBG2	Q8kbg2 chlorobium
22	47	44.3	293	16	O25975	O25975 helicobacte
23	47	44.3	293	16	Q9ZJI2	Q9zji2 helicobacte
24	47	44.3	390	16	Q98CK1	Q98ck1 rhizobium l
25	47	44.3	661	10	Q8S8J8	Q8s8j8 arabidopsis
26	46.5	43.9	242	16	Q7WCX1	Q7wcx1 bordetella
27	46.5	43.9	298	16	Q7W5D6	Q7w5d6 bordetella
28	46.5	43.9	298	16	Q7VTI2	Q7vti2 bordetella
29	46.5	43.9	563	16	Q9X8S9	Q9x8s9 streptomyce
30	46	43.4	140	16	Q8P5G3	Q8p5g3 xanthomonas
31	46	43.4	143	4	Q8NAW2	Q8naw2 homo sapien
32	46	43.4	192	15	Q90CX5	Q90cx5 human immun
33	46	43.4	192	15	Q8UTJ5	Q8utj5 human immun
34	46	43.4	192	15	O70887	O70887 human immun
35	46	43.4	274	16	Q9PPC9	Q9ppc9 campylobact
36	46	43.4	300	16	O34990	O34990 bacillus su
37	46	43.4	462	16	Q9KZB0	Q9kzb0 streptomyce
38	46	43.4	753	16	Q89T31	Q89t31 bradyrhizob
39	45	42.5	85	6	O77786	O77786 canis famil
40	45	42.5	192	15	Q90ML0	Q90ml0 human immun
41	45	42.5	192	15	Q994N4	Q994n4 human immun
42	45	42.5	192	15	Q9IW51	Q9iw51 human immun
43	45	42.5	192	15	Q9WQH9	Q9wqh9 human immun
44	45	42.5	192	15	Q9WP98	Q9wp98 human immun
45	45	42.5	192	15	Q99BN6	Q99bn6 human immun

ALIGNMENTS

RESULT 1
Q8K408
ID Q8K408 PRELIMINARY; PRT; 41 AA.
AC Q8K408;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Truncated macrophage colony stimulating factor.
GN CSF1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEW.tl;
RX MEDLINE=22069908; PubMed=12074592;
RA Dobbins D.E., Sood R., Hashiramoto A., Hansen C.T., Wilder R.L.,
RA Remmers E.F.;
RT "Mutation of macrophage colony stimulating factor (Csf1) causes
RT osteopetrosis in the tl rat.";
RL Biochem. Biophys. Res. Commun. 294:1114-1120(2002).
DR EMBL; AF514357; AAM54137.1; -.
SQ SEQUENCE 41 AA; 4178 MW; 1D342C19BD18AA41 CRC64;

Query Match 60.4%; Score 64; DB 11; Length 41;
Best Local Similarity 70.0%; Pred. No. 0.014;
Matches 14; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy          1 MAGLPAVVGLSPGEQEYHRG 20
              | ||||  |||| ||| ||
Db         18 MDGLPAAAGLSPREQECCRG 37
```

RESULT 2
Q8PE55
ID Q8PE55 PRELIMINARY; PRT; 150 AA.
AC Q8PE55;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)

```
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Deoxycytidylate deaminase.
GN   XCC0126.
OS   Xanthomonas campestris (pv. campestris).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xanthomonas.
OX   NCBI_TaxID=340;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33913 / NCPPB 528;
RX   MEDLINE=22022145; PubMed=12024217;
RA   da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA   Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA   Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA   Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA   Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA   Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA   Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA   Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA   Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA   Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA   Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA   Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA   Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA   Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA   Setubal J.C., Kitajima J.P.;
RT   "Comparison of the genomes of two Xanthomonas pathogens with differing
RT   host specificities.";
RL   Nature 417:459-463(2002).
DR   EMBL; AE012107; AAM39445.1; -.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   InterPro; IPR002125; dCMP/cyt_deam.
DR   Pfam; PF00383; dCMP_cyt_deam; 1.
KW   Complete proteome.
SQ   SEQUENCE   150 AA;  16201 MW;  63D9C17D44DC8B43 CRC64;

  Query Match             49.1%;  Score 52;  DB 16;  Length 150;
  Best Local Similarity   58.8%;  Pred. No. 4.1;
  Matches   10;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 AGLPAVVGLSPGEQEYH 18
              ||:  || |:||| | |
Db        104 AGIKRVVALAPGESESH 120


RESULT 3
Q89Y25
ID   Q89Y25      PRELIMINARY;      PRT;   242 AA.
AC   Q89Y25;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Bll0130 protein.
GN   BLL0130.
OS   Bradyrhizobium japonicum.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Bradyrhizobiaceae; Bradyrhizobium.
OX   NCBI_TaxID=375;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=USDA 110;
RX   MEDLINE=22484998; PubMed=12597275;
RA   Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA   Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA   Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA   Tabata S.;
RT   "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT   Bradyrhizobium japonicum USDA110.";
RL   DNA Res. 9:189-197(2002).
DR   EMBL; AP005935; BAC45395.1; -.
DR   InterPro; IPR005493; Methyltransf_6.
DR   Pfam; PF03737; Methyltransf_6; 1.
KW   Complete proteome.
SQ   SEQUENCE   242 AA;  25216 MW;  588F9C0B9396414B CRC64;

  Query Match             47.2%;  Score 50;  DB 16;  Length 242;
  Best Local Similarity   55.0%;  Pred. No. 14;
  Matches   11;  Conservative    1;  Mismatches    4;  Indels    4;  Gaps    1;

Qy          1 MAGLPAVVGLSPGEQEYHRG 20
              :| ||  ||| |    ||||
Db         63 LARLPGAVGLKP----YHRG 78


RESULT 4
Q9KQK6
ID   Q9KQK6      PRELIMINARY;      PRT;   277 AA.
AC   Q9KQK6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Formyltetrahydrofolate deformylase.
GN   VC1992.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=El Tor N16961 / Serotype O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004274; AAF95140.1; -.
DR   PIR; F82130; F82130.
DR   HSSP; P08179; 2GAR.
DR   TIGR; VC1992; -.
DR   GO; GO:0016597; F:amino acid binding; IEA.
DR   GO; GO:0008864; F:formyltetrahydrofolate deformylase activity; IEA.
DR   GO; GO:0016742; F:hydroxymethyl-, formyl- and related transfe. . .; IEA.
DR   GO; GO:0006189; P:'de novo' IMP biosynthesis; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   InterPro; IPR002912; ACT.
DR   InterPro; IPR002376; formyl_transf.
DR   InterPro; IPR004810; PurU.
DR   Pfam; PF01842; ACT; 1.
DR   Pfam; PF00551; formyl_transf; 1.
DR   PRINTS; PR01575; FFH4HYDRLASE.
DR   TIGRFAMs; TIGR00655; PurU; 1.
KW   Complete proteome.
SQ   SEQUENCE   277 AA;  31373 MW;  A703491654753DC6 CRC64;

  Query Match             47.2%;  Score 50;  DB 16;  Length 277;
  Best Local Similarity   52.9%;  Pred. No. 16;
  Matches    9;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :|  | :| | ||
Db        190 LPAFIGAKPYQQAYERG 206


RESULT 5
Q7VFI5
ID   Q7VFI5      PRELIMINARY;      PRT;   284 AA.
AC   Q7VFI5;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
```

```
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Formyltetrahydrofolate deformylase PurU (EC 3.5.1.10).
GN   PURU OR HH1691.
OS   Helicobacter hepaticus.
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=32025;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 51449 / 3B1;
RX   MEDLINE=22709201; PubMed=12810954;
RA   Suerbaum S., Josenhans C., Sterzenbach T., Drescher B., Brandt P.,
RA   Bell M., Droege M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,
RA   Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,
RA   Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.;
RT   "The complete genome sequence of the carcinogenic bacterium
RT   Helicobacter hepaticus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906(2003).
DR   EMBL; AE017149; AAP78288.1; -.
KW   Hydrolase; Complete proteome.
SQ   SEQUENCE   284 AA;  32660 MW;  AAED2DC5086236E3 CRC64;

  Query Match             47.2%;  Score 50;  DB 16;  Length 284;
  Best Local Similarity   52.9%;  Pred. No. 17;
  Matches    9;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :| :| :| | ||
Db        197 LPAFIGANPYKQAYERG 213


RESULT 6
Q9RRL5
ID   Q9RRL5      PRELIMINARY;      PRT;   381 AA.
AC   Q9RRL5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cytochrome P450.
GN   DR2473.
OS   Deinococcus radiodurans.
OC   Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AE002076; AAF12016.1; -.
DR   PIR; F75270; F75270.
DR   TIGR; DR2473; -.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase; Complete proteome.
SQ   SEQUENCE   381 AA;  41940 MW;  F191EA69F1797B53 CRC64;

  Query Match             47.2%;  Score 50;  DB 16;  Length 381;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 GLPAVVGLSP 12
              ||||||||||
Db         51 GLPAVVGLSP 60


RESULT 7
Q87RD3
ID   Q87RD3      PRELIMINARY;      PRT;   277 AA.
AC   Q87RD3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Formyltetrahydrofolate deformylase.
GN   VP0864.
OS   Vibrio parahaemolyticus.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=670;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIMD 2210633 / Serotype O3:K6;
RX   MEDLINE=22508454; PubMed=12620739;
RA   Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA   Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA   Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT   "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT   distinct from that of V. cholerae.";
RL   Lancet 361:743-749(2003).
DR   EMBL; AP005075; BAC59127.1; -.
DR   GO; GO:0016597; F:amino acid binding; IEA.
DR   GO; GO:0008864; F:formyltetrahydrofolate deformylase activity; IEA.
DR   GO; GO:0016742; F:hydroxymethyl-, formyl- and related transfe. . .; IEA.
DR   GO; GO:0006189; P:'de novo' IMP biosynthesis; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   InterPro; IPR002912; ACT.
DR   InterPro; IPR002376; formyl_transf.
DR   InterPro; IPR004810; PurU.
DR   Pfam; PF01842; ACT; 1.
DR   Pfam; PF00551; formyl_transf; 1.
DR   PRINTS; PR01575; FFH4HYDRLASE.
KW   Complete proteome.
SQ   SEQUENCE   277 AA;  31620 MW;  D5F0712E1537CF4C CRC64;

  Query Match             46.2%;  Score 49;  DB 16;  Length 277;
  Best Local Similarity   52.9%;  Pred. No. 23;
  Matches    9;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :|  | :| | ||
Db        190 LPAFIGAKPYQQAYDRG 206


RESULT 8
O93321
ID   O93321      PRELIMINARY;      PRT;  4823 AA.
AC   O93321;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   All-1 related protein.
GN   ALR.
OS   Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetradontoidea; Tetraodontidae; Takifugu.
OX   NCBI_TaxID=31033;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99177347; PubMed=10077531;
```

```
RA   Gellner K., Brenner S.;
RT   "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
RT   rubripes.";
RL   Genome Res. 9:251-258(1999).
CC   -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
DR   EMBL; AF056116; AAC34383.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR003889; FYrich_C.
DR   InterPro; IPR003888; FYrich_N.
DR   InterPro; IPR000910; HMG_12_box.
DR   InterPro; IPR003616; PostSET.
DR   InterPro; IPR001214; SET.
DR   InterPro; IPR001965; Znf_PHD.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00628; PHD; 5.
DR   Pfam; PF00856; SET; 1.
DR   SMART; SM00542; FYRC; 1.
DR   SMART; SM00541; FYRN; 1.
DR   SMART; SM00398; HMG; 1.
DR   SMART; SM00249; PHD; 8.
DR   SMART; SM00508; PostSET; 1.
DR   SMART; SM00184; RING; 4.
DR   SMART; SM00317; SET; 1.
DR   PROSITE; PS50868; POST_SET; 1.
DR   PROSITE; PS50280; SET; 1.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
SQ   SEQUENCE   4823 AA;  526260 MW;  BD0C5F4EAD0F9CC7 CRC64;

  Query Match             46.2%;  Score 49;  DB 13;  Length 4823;
  Best Local Similarity   66.7%;  Pred. No. 5.2e+02;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          5 PAVVGLSPGEQEYHR 19
              ||: |||| | | ||
Db       1900 PALGGLSPSELEKHR 1914


RESULT 9
Q8UTQ8
ID   Q8UTQ8      PRELIMINARY;      PRT;   192 AA.
AC   Q8UTQ8;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Vif protein (Virion infectivity factor) (SOR protein).
GN   VIF.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=98BWMC14.a3;
RA   Novitsky V.A., Smith U.R., Gilbert P., McLane M.F., Chigwedere P.,
RA   Williamson C., Ndung'u T., Klein I., Chang S.-Y., Peter T., Thior I.,
RA   Foley B.T., Gaolekwe S., Rybak N., Gaseitsiwe S., Vannberg F.,
RA   Marlink R., Lee T.-H., Essex M.;
RT   "HIV-1 subtype C molecular phylogeny: consensus sequence for an AIDS
RT   vaccine design.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: DETERMINES VIRUS INFECTIVITY (BY SIMILARITY).
DR   EMBL; AF443078; AAL34591.1; -.
DR   GO; GO:0019058; P:viral infectious cycle; IEA.
DR   InterPro; IPR000475; Viral_infect.
DR   Pfam; PF00559; Vif; 1.
DR   PRINTS; PR00349; VIRIONINFFCT.
DR   ProDom; PD000063; Viral_infect; 1.
KW   AIDS.
SQ   SEQUENCE   192 AA;  22624 MW;  5E393F973B1F8C4A CRC64;

  Query Match             45.3%;  Score 48;  DB 15;  Length 192;
  Best Local Similarity   66.7%;  Pred. No. 22;
  Matches    8;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          9 GLSPGEQEYHRG 20
              || |||:|:| |
Db         71 GLQPGEREWHLG 82


RESULT 10
Q90D29
ID   Q90D29      PRELIMINARY;      PRT;   192 AA.
AC   Q90D29;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Vif protein (Virion infectivity factor) (SOR protein).
GN   VIF.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=97TZ01;
RX   MEDLINE=21395692; PubMed=11504977;
RA   Hoelscher M., Kim B., Maboko L., Mhalu F., von Sonnenburg F.,
RA   Birx D.L., McCutchan F.E.,
RA   the UNAIDS Network for HIV Isolation Characterization.;
RT   "High proportion of unrelated HIV-1 intersubtype recombinants in the
RT   Mbeya region of southwest Tanzania.";
RL   AIDS 15:1461-1470(2001).
CC   -!- FUNCTION: DETERMINES VIRUS INFECTIVITY (BY SIMILARITY).
DR   EMBL; AF361871; AAK94212.1; -.
DR   GO; GO:0019058; P:viral infectious cycle; IEA.
DR   InterPro; IPR000475; Viral_infect.
DR   Pfam; PF00559; Vif; 1.
DR   PRINTS; PR00349; VIRIONINFFCT.
DR   ProDom; PD000063; Viral_infect; 1.
KW   AIDS.
SQ   SEQUENCE   192 AA;  22674 MW;  11B799C5DEA99F77 CRC64;

  Query Match             45.3%;  Score 48;  DB 15;  Length 192;
  Best Local Similarity   66.7%;  Pred. No. 22;
  Matches    8;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          9 GLSPGEQEYHRG 20
              || |||:|:| |
Db         71 GLQPGEREWHLG 82


RESULT 11
Q9CMF6
ID   Q9CMF6      PRELIMINARY;      PRT;   278 AA.
AC   Q9CMF6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   PurU.
GN   PURU OR PM0873.
OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Pasteurella.
OX   NCBI_TaxID=747;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pm70;
RX   MEDLINE=21145866; PubMed=11248100;
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR   EMBL; AE006126; AAK02957.1; -.
DR   HSSP; P08179; 1GRC.
DR   GO; GO:0016597; F:amino acid binding; IEA.
```

```
DR   GO; GO:0008864; F:formyltetrahydrofolate deformylase activity; IEA.
DR   GO; GO:0016742; F:hydroxymethyl-, formyl- and related transfe. . .; IEA.
DR   GO; GO:0006189; P:'de novo' IMP biosynthesis; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   InterPro; IPR002912; ACT.
DR   InterPro; IPR002376; formyl_transf.
DR   InterPro; IPR004810; PurU.
DR   Pfam; PF01842; ACT; 1.
DR   Pfam; PF00551; formyl_transf; 1.
DR   PRINTS; PR01575; FFH4HYDRLASE.
DR   TIGRFAMs; TIGR00655; PurU; 1.
KW   Complete proteome.
SQ   SEQUENCE   278 AA;  32086 MW;  F303E499DFFA0B70 CRC64;

  Query Match             45.3%;  Score 48;  DB 16;  Length 278;
  Best Local Similarity   52.9%;  Pred. No. 33;
  Matches    9;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :|  |  | | ||
Db        191 LPAFIGAKPYHQAYERG 207


RESULT 12
Q8EMZ7
ID   Q8EMZ7      PRELIMINARY;      PRT;   300 AA.
AC   Q8EMZ7;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Formyltetrahydrofolate deformylase (EC 3.5.1.10).
GN   OB2693.
OS   Oceanobacillus iheyensis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX   NCBI_TaxID=182710;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX   MEDLINE=22220767; PubMed=12235376;
RA   Takami H., Takaki Y., Uchiyama I.;
RT   "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT   Ridge and its unexpected adaptive capabilities to extreme
RT   environments.";
RL   Nucleic Acids Res. 30:3927-3935(2002).
DR   EMBL; AP004602; BAC14649.1; -.
DR   GO; GO:0016597; F:amino acid binding; IEA.
DR   GO; GO:0008864; F:formyltetrahydrofolate deformylase activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0016742; F:hydroxymethyl-, formyl- and related transfe. . .; IEA.
DR   GO; GO:0006189; P:'de novo' IMP biosynthesis; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   InterPro; IPR002912; ACT.
DR   InterPro; IPR002376; formyl_transf.
DR   InterPro; IPR004810; PurU.
DR   Pfam; PF01842; ACT; 1.
DR   Pfam; PF00551; formyl_transf; 1.
DR   PRINTS; PR01575; FFH4HYDRLASE.
DR   TIGRFAMs; TIGR00655; PurU; 1.
KW   Hydrolase; Complete proteome.
SQ   SEQUENCE   300 AA;  34894 MW;  D44D0FE0CC596BF2 CRC64;

  Query Match             45.3%;  Score 48;  DB 16;  Length 300;
  Best Local Similarity   52.9%;  Pred. No. 36;
  Matches    9;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :|  | |: | ||
Db        212 LPAFIGAKPYERAYDRG 228


RESULT 13
Q7XNE4
ID   Q7XNE4      PRELIMINARY;      PRT;   485 AA.
AC   Q7XNE4;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   OSJNBa0009P12.18 protein.
GN   OSJNBA0009P12.18.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA   Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
RA   Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA   Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA   Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA   Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA   Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA   Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA   Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL662952; CAE04133.1; -.
SQ   SEQUENCE   485 AA;  52334 MW;  F7685952D5538456 CRC64;

  Query Match             45.3%;  Score 48;  DB 10;  Length 485;
  Best Local Similarity   60.0%;  Pred. No. 61;
  Matches    9;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYH 18
              |||:  ||||:  ||
Db        317 LPALSKLSPGQAAYH 331


RESULT 14
Q994Q2
ID   Q994Q2      PRELIMINARY;      PRT;   192 AA.
AC   Q994Q2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Vif protein (Virion infectivity factor) (SOR protein).
GN   VIF.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=96ZM751;
RX   MEDLINE=21094715; PubMed=11177395;
RA   Rodenburg C.M., Li Y., Trask S.A., Chen Y., Decker J., Robertson D.L.,
RA   Kalish M.L., Shaw G.M., Allen S., Hahn B.H., Gao F.;
RT   "Near full-length clones and reference sequences for subtype C
RT   isolates for HIV type 1 from three different continents.";
RL   AIDS Res. Hum. Retroviruses 17:161-168(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=96ZM751;
RA   Rodenburg C.M., Li Y., Trask S.A., Chen Y., Decker J., Robertson D.L.,
RA   Allen S., Shaw G.M., Hahn B.H., Gao F.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: DETERMINES VIRUS INFECTIVITY (BY SIMILARITY).
DR   EMBL; AF286225; AAK30974.1; -.
DR   GO; GO:0019058; P:viral infectious cycle; IEA.
DR   InterPro; IPR000475; Viral_infect.
DR   Pfam; PF00559; Vif; 1.
DR   PRINTS; PR00349; VIRIONINFFCT.
DR   ProDom; PD000063; Viral_infect; 1.
KW   AIDS.
SQ   SEQUENCE   192 AA;  22727 MW;  55E01D4BBCD93DC6 CRC64;

  Query Match             44.3%;  Score 47;  DB 15;  Length 192;
```

```
  Best Local Similarity   66.7%;  Pred. No. 32;
  Matches    8;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          9 GLSPGEQEYHRG 20
              || |||:|:| |
Db         71 GLHPGEREWHLG 82


RESULT 15
Q9QSR2
ID   Q9QSR2      PRELIMINARY;      PRT;   192 AA.
AC   Q9QSR2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Vif protein (Virion infectivity factor) (SOR protein).
GN   VIF.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VI850;
RX   MEDLINE=20192166; PubMed=10725202;
RA   Laukkanen T., Carr J.K., Janssens W., Liitsola K., Gotte D.,
RA   McCutchan F.E., Op de Coul E., Cornelissen M., Heyndrickx L.,
RA   van der Groen G., Salminen M.O.;
RT   "Virtually full-length subtype F and F/D recombinant HIV-1 from Africa
RT   and South America.";
RL   Virology 269:95-104(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VI850;
RA   Carr J.K., Kim B., Sanders-Buell E., Salminen M.O., Alaeus A.,
RA   Albert J.A., Birx D.L., McCutchan F.E.;
RT   "HIV-1 isolate VI850 from Zaire, complete genome.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: DETERMINES VIRUS INFECTIVITY (BY SIMILARITY).
DR   EMBL; AF077336; AAD46089.1; -.
DR   GO; GO:0019058; P:viral infectious cycle; IEA.
DR   InterPro; IPR000475; Viral_infect.
DR   Pfam; PF00559; Vif; 1.
DR   PRINTS; PR00349; VIRIONINFFCT.
DR   ProDom; PD000063; Viral_infect; 1.
KW   AIDS.
SQ   SEQUENCE   192 AA;  22629 MW;  F165805BFCDA4427 CRC64;

  Query Match             44.3%;  Score 47;  DB 15;  Length 192;
  Best Local Similarity   66.7%;  Pred. No. 32;
  Matches    8;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          9 GLSPGEQEYHRG 20
              || |||:|:| |
Db         71 GLHPGEREWHLG 82


Search completed: May  7, 2004, 12:37:52
Job time : 39.5667 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:19:40 ; Search time 55.8 Seconds
(without alignments)
101.272 Million cell updates/sec

Title: US-09-786-214A-9
Perfect score: 106
Sequence: 1 MAGLPAVVGLSPGEQEYHRG 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	106	100.0	20	3	AAY84265	Aay84265 Truncated
2	106	100.0	25	3	AAY84264	Aay84264 Peptide o
3	75	70.8	15	3	AAY84269	Aay84269 Peptide d
4	72	67.9	14	3	AAY84266	Aay84266 Peptide d
5	68	64.2	13	3	AAY84267	Aay84267 Peptide d
6	65	61.3	13	3	AAY84268	Aay84268 Peptide d
7	50	47.2	277	6	ABU49441	Abu49441 Protein e
8	49	46.2	191	3	AAB69296	Aab69296 HIV-1 non
9	49	46.2	278	6	ABU30626	Abu30626 Protein e
10	48	45.3	278	6	ABU39195	Abu39195 Protein e
11	48	45.3	287	6	ABU17217	Abu17217 Protein e
12	48	45.3	303	6	ADA36829	Ada36829 Acinetoba
13	47	44.3	151	5	ABU50765	Abu50765 Helicobac
14	47	44.3	155	5	ABU51750	Abu51750 Helicobac
15	47	44.3	192	3	AAB69298	Aab69298 HIV-1 non
16	47	44.3	238	6	ABU27483	Abu27483 Protein e
17	47	44.3	278	6	ABU44771	Abu44771 Protein e
18	47	44.3	280	6	ABU48054	Abu48054 Protein e
19	47	44.3	280	6	ABU31449	Abu31449 Protein e
20	47	44.3	280	6	ABU15054	Abu15054 Protein e
21	47	44.3	280	6	ABU47098	Abu47098 Protein e
22	47	44.3	282	6	ABU41187	Abu41187 Protein e
23	47	44.3	283	6	ABM69037	Abm69037 Photorhab
24	47	44.3	293	2	AAW98485	Aaw98485 H. pylori
25	47	44.3	293	6	ABU31073	Abu31073 Protein e
26	46	43.4	143	5	AAO21677	Aao21677 Human sec
27	46	43.4	143	7	ADB64633	Adb64633 Human pro
28	46	43.4	187	6	ABU21296	Abu21296 Protein e
29	46	43.4	192	3	AAB69290	Aab69290 HIV-1 non
30	46	43.4	234	4	AAB36208	Aab36208 Human imm
31	46	43.4	274	6	ABU26498	Abu26498 Protein e
32	46	43.4	300	2	AAY16108	Aay16108 A formate
33	45	42.5	116	4	ABG24038	Abg24038 Novel hum
34	45	42.5	287	7	ADB80057	Adb80057 Mycobacte
35	45	42.5	334	4	ABB11182	Abb11182 Human gua
36	45	42.5	334	4	AAM79690	Aam79690 Human pro
37	45	42.5	403	4	AAM40847	Aam40847 Human pol
38	45	42.5	456	4	AAM78706	Aam78706 Human pro
39	45	42.5	456	4	AAM39061	Aam39061 Human pol
40	44	41.5	506	4	ABB71810	Abb71810 Drosophil
41	44	41.5	565	4	AAB76817	Aab76817 Corynebac
42	44	41.5	579	4	AAG90917	Aag90917 C glutami
43	44	41.5	940	4	AAU28194	Aau28194 Novel hum
44	44	41.5	968	4	AAU28382	Aau28382 Novel hum
45	44	41.5	1663	6	ABO53095	Abo53095 Human put

ALIGNMENTS

RESULT 1
AAY84265
ID AAY84265 standard; peptide; 20 AA.
XX
AC AAY84265;
XX
DT 12-JUL-2000 (first entry)
XX
DE Truncated macrophage colony stimulating factor tumour antigen.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
DR N-PSDB; AAZ99675.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Claim 3; Page 64; 74pp; English.
XX
CC The present sequence represents a truncated tumour rejection antigen
CC precursor, and is encoded by a truncated alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying

```
CC   functional variants and mimetics
XX
SQ   Sequence 20 AA;

  Query Match             100.0%;  Score 106;  DB 3;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-09;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAGLPAVVGLSPGEQEYHRG 20
              ||||||||||||||||||||
Db          1 MAGLPAVVGLSPGEQEYHRG 20


RESULT 2
AAY84264
ID   AAY84264 standard; peptide; 25 AA.
XX
AC   AAY84264;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Peptide of alternate reading frame of macrophage colony stimulating gene.
XX
KW   Renal cell carcinoma; antigen; cytotoxic T lymphocyte;
KW   tumour rejection antigen; macrophage colony stimulating gene;
KW   macrophage-colony stimulating factor; antigen presenting cell;
KW   human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS   Homo sapiens.
XX
PN   WO200013699-A1.
XX
PD   16-MAR-2000.
XX
PF   03-SEP-1999;   99WO-US020344.
XX
PR   04-SEP-1998;   98US-0099077P.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Probst-Kepper M,  Van Den Eynde B,  Boon-Falleur T;
XX
DR   WPI; 2000-256859/22.
DR   N-PSDB; AAZ99672.
XX
PT   Isolated polypeptide used to treat subjects having a disorder
PT   characterized by expression of alternative open reading frame macrophage-
PT   colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS   Claim 1; Page 64; 74pp; English.
XX
CC   The present sequence represents a tumour rejection antigen precursor, and
CC   is encoded by an alternative open reading frame (ORF) of human macrophage
CC   colony stimulating gene. Peptides derived from the alternative ORF of
CC   macrophage-colony stimulating factor, when presented by an antigen
CC   presenting cell having a human leukocyte antigen (HLA) class I molecule,
CC   effectively induce the activation and proliferation of CD8+ cytotoxic T
CC   lymphocytes. Polypeptide and nucleic acids derived from the alternate ORF
CC   of macrophage-colony stimulating factor are useful for enriching
CC   selectively a population of T lymphocytes with CD8+ T lymphocytes. They
CC   are also useful for diagnosing a disorder characterized by expression of
CC   the polypeptide, and for identifying functional variants and mimetics
XX
SQ   Sequence 25 AA;

  Query Match             100.0%;  Score 106;  DB 3;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-08;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAGLPAVVGLSPGEQEYHRG 20
              ||||||||||||||||||||
Db          1 MAGLPAVVGLSPGEQEYHRG 20


RESULT 3
AAY84269
ID   AAY84269 standard; peptide; 15 AA.
XX
AC   AAY84269;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW   tumour rejection antigen; macrophage colony stimulating gene;
KW   macrophage-colony stimulating factor; antigen presenting cell;
KW   human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO200013699-A1.
XX
PD   16-MAR-2000.
XX
PF   03-SEP-1999;   99WO-US020344.
XX
PR   04-SEP-1998;   98US-0099077P.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Probst-Kepper M,  Van Den Eynde B,  Boon-Falleur T;
XX
DR   WPI; 2000-256859/22.
XX
PT   Isolated polypeptide used to treat subjects having a disorder
PT   characterized by expression of alternative open reading frame macrophage-
PT   colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS   Example 2; Page 40; 74pp; English.
XX
CC   The present sequence represents a peptide which is derived from a tumour
CC   rejection antigen precursor encoded by an alternative open reading frame
CC   (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC   the alternative ORF of macrophage-colony stimulating factor, when
CC   presented by an antigen presenting cell having a human leukocyte antigen
CC   (HLA) class I molecule, effectively induce the activation and
CC   proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC   acids derived from the alternate ORF of macrophage-colony stimulating
CC   factor are useful for enriching selectively a population of T lymphocytes
CC   with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC   characterized by expression of the polypeptide, and for identifying
CC   functional variants and mimetics
XX
SQ   Sequence 15 AA;

  Query Match             70.8%;  Score 75;  DB 3;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.00026;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 AGLPAVVGLSPGEQE 16
              |||||||||||||||
Db          1 AGLPAVVGLSPGEQE 15


RESULT 4
AAY84266
ID   AAY84266 standard; peptide; 14 AA.
XX
AC   AAY84266;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW   tumour rejection antigen; macrophage colony stimulating gene;
```

KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Claim 2; Page 39; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics
XX
SQ Sequence 14 AA;

Query Match 67.9%; Score 72; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00069;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 LPAVVGLSPGEQEY 17
||||||||||||||
Db 1 LPAVVGLSPGEQEY 14

RESULT 5
AAY84267
ID AAY84267 standard; peptide; 13 AA.
XX
AC AAY84267;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Example 2; Page 40; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics
XX
SQ Sequence 13 AA;

Query Match 64.2%; Score 68; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 PAVVGLSPGEQEY 17
|||||||||||||
Db 1 PAVVGLSPGEQEY 13

RESULT 6
AAY84268
ID AAY84268 standard; peptide; 13 AA.
XX
AC AAY84268;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX

```
PS   Example 2; Page 40; 74pp; English.
XX
CC   The present sequence represents a peptide which is derived from a tumour
CC   rejection antigen precursor encoded by an alternative open reading frame
CC   (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC   the alternative ORF of macrophage-colony stimulating factor, when
CC   presented by an antigen presenting cell having a human leukocyte antigen
CC   (HLA) class I molecule, effectively induce the activation and
CC   proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC   acids derived from the alternate ORF of macrophage-colony stimulating
CC   factor are useful for enriching selectively a population of T lymphocytes
CC   with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC   characterized by expression of the polypeptide, and for identifying
CC   functional variants and mimetics
XX
SQ   Sequence 13 AA;

  Query Match             61.3%;  Score 65;  DB 3;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 0.007;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQE 16
              |||||||||||||
Db          1 LPAVVGLSPGEQE 13


RESULT 7
ABU49441
ID   ABU49441 standard; protein; 277 AA.
XX
AC   ABU49441;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #34968.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Vibrio cholerae.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA53311.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 77365; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 277 AA;

  Query Match             47.2%;  Score 50;  DB 6;  Length 277;
  Best Local Similarity   52.9%;  Pred. No. 33;
  Matches    9;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :|  | :| | ||
Db        190 LPAFIGAKPYQQAYERG 206


RESULT 8
AAB69296
ID   AAB69296 standard; protein; 191 AA.
XX
AC   AAB69296;
XX
DT   12-SEP-2003  (revised)
DT   20-APR-2001  (first entry)
XX
DE   HIV-1 non-subtype B clone 94CY032-3 vif protein.
XX
KW   HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW   vif; vpr; tat; rev; nef; vaccine.
XX
OS   Human immunodeficiency virus 1.
XX
PN   WO200026416-A1.
XX
PD   11-MAY-2000.
XX
PF   25-OCT-1999;  99WO-US024837.
XX
PR   02-NOV-1998;  98US-00184418.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Hahn BH,  Shaw GM,  Gao F;
XX
DR   WPI; 2000-365651/31.
XX
PT   Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT   type 1 useful for detecting and treating AIDS comprises a specific
PT   nucleotide sequence.
XX
PS   Claim 41; Fig 16; 131pp; English.
XX
CC   The present in invention provides the protein and coding sequences for a
```

CC number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC isolates. The sequences shown include the near full-length coding
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC rev and nef proteins. These can be used to detect the presence of HIV-1
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC These antibodies can be used in vaccines to prevent and treat HIV
CC infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ Sequence 191 AA;

Query Match 46.2%; Score 49; DB 3; Length 191;
Best Local Similarity 66.7%; Pred. No. 31;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 9 GLSPGEQEYHRG 20
|| ||||::| |
Db 70 GLQPGEQDWHLG 81

RESULT 9
ABU30626
ID ABU30626 standard; protein; 278 AA.
XX
AC ABU30626;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #16153.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Haemophilus influenzae.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA34496.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 58550; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 278 AA;

Query Match 46.2%; Score 49; DB 6; Length 278;
Best Local Similarity 52.9%; Pred. No. 46;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 4 LPAVVGLSPGEQEYHRG 20
||| :| | :| | ||
Db 191 LPAFIGAKPYQQAYKRG 207

RESULT 10
ABU39195
ID ABU39195 standard; protein; 278 AA.
XX
AC ABU39195;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #24722.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Pasteurella multocida.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA43065.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 67119; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 278 AA;

Query Match 45.3%; Score 48; DB 6; Length 278;
Best Local Similarity 52.9%; Pred. No. 65;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

```
Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :|  |  | | ||
Db        191 LPAFIGAKPYHQAYERG 207
```

RESULT 11
ABU17217
ID ABU17217 standard; protein; 287 AA.
XX
AC ABU17217;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #2744.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Acinetobacter baumannii.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA21087.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 45141; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 287 AA;

Query Match 45.3%; Score 48; DB 6; Length 287;
Best Local Similarity 52.9%; Pred. No. 68;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```
Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| || :| :| | :|
Db        200 LPAFVGANPYKQAYEKG 216
```

RESULT 12
ADA36829
ID ADA36829 standard; protein; 303 AA.
XX
AC ADA36829;
XX
DT 20-NOV-2003 (first entry)
XX
DE Acinetobacter baumannii protein #3990.
XX
KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
KW plant biocontrol agent.
XX
OS Acinetobacter baumannii.
XX
PN US6562958-B1.
XX
PD 13-MAY-2003.
XX
PF 04-JUN-1999; 99US-00328352.
XX
PR 09-JUN-1998; 98US-0088701P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.

XX
PI Breton G, Bush D;
XX
DR WPI; 2003-576092/54.
DR N-PSDB; ADA32703.
XX
PT New Acinetobacter baumanii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX
PS Example; SEQ ID NO 8116; 328pp; English.
XX
CC The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents the amino acid sequence of an A.
CC baumannii protein.
XX
SQ Sequence 303 AA;

Query Match 45.3%; Score 48; DB 6; Length 303;
Best Local Similarity 52.9%; Pred. No. 72;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 4 LPAVVGLSPGEQEYHRG 20
||| || :| :| | :|
Db 216 LPAFVGANPYKQAYEKG 232

RESULT 13
ABU50765
ID ABU50765 standard; protein; 151 AA.
XX
AC ABU50765;
XX
DT 07-MAY-2003 (first entry)
XX
DE Helicobacter pylori selected interacting domain (SID) protein #108.
XX
KW Protein-protein interaction; ulcer; selected interacting domain; SID.
XX
OS Helicobacter pylori.
XX
PN WO200266501-A2.
XX
PD 29-AUG-2002.
XX
PF 28-DEC-2001; 2001WO-EP015428.
XX
PR 02-JAN-2001; 2001US-0259302P.
XX
PA (HYBR-) HYBRIGENICS.
PA (INSP) INST PASTEUR.
XX
PI Legrain P, Rain J, Colland F, De Reuse H, Labigne A;
XX
DR WPI; 2002-674910/72.
DR N-PSDB; ABX65509.
XX
PT New complexes of protein-protein interactions in Helicobacter pylori,
PT useful for identifying modulating compounds for treating or preventing
PT ulcers in mammals.
XX
PS Claim 6; Page 111; 642pp; English.
XX
CC The invention describes a complex of protein-protein interactions in
CC Helicobacter pylori selected from 421 complexes given in the
CC specification. The complex of protein-protein interactions are useful for
CC screening for agents which modulate the interaction of proteins.
CC Modulating compounds which binds to a targeted bacterial protein may be
CC used for treating or preventing ulcers in a human or animal. This is the
CC amino acid sequence of a selected interacting domain (SID), identified
CC via protein-protein interactions. Note: Where the patent number printed
CC at the top of the pages in the specification has obscured areas of
CC protein sequence, the indexer has replaced the residue with an X to
CC represent an illegible residue
XX
SQ Sequence 151 AA;

Query Match 44.3%; Score 47; DB 5; Length 151;
Best Local Similarity 47.1%; Pred. No. 48;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 4 LPAVVGLSPGEQEYHRG 20
||| :| :| :| : ||
Db 79 LPAFIGANPYQQAFERG 95

RESULT 14
ABU51750
ID ABU51750 standard; protein; 155 AA.
XX
AC ABU51750;
XX
DT 07-MAY-2003 (first entry)
XX
DE Helicobacter pylori selected interacting domain (SID) protein #1094.
XX
KW Protein-protein interaction; ulcer; selected interacting domain; SID.
XX
OS Helicobacter pylori.
XX
PN WO200266501-A2.
XX
PD 29-AUG-2002.
XX
PF 28-DEC-2001; 2001WO-EP015428.
XX
PR 02-JAN-2001; 2001US-0259302P.
XX
PA (HYBR-) HYBRIGENICS.
PA (INSP) INST PASTEUR.
XX
PI Legrain P, Rain J, Colland F, De Reuse H, Labigne A;
XX
DR WPI; 2002-674910/72.
DR N-PSDB; ABX66495.
XX
PT New complexes of protein-protein interactions in Helicobacter pylori,
PT useful for identifying modulating compounds for treating or preventing
PT ulcers in mammals.
XX
PS Claim 6; Page 354; 642pp; English.
XX
CC The invention describes a complex of protein-protein interactions in
CC Helicobacter pylori selected from 421 complexes given in the
CC specification. The complex of protein-protein interactions are useful for
CC screening for agents which modulate the interaction of proteins.
CC Modulating compounds which binds to a targeted bacterial protein may be
CC used for treating or preventing ulcers in a human or animal. This is the
CC amino acid sequence of a selected interacting domain (SID), identified
CC via protein-protein interactions. Note: Where the patent number printed
CC at the top of the pages in the specification has obscured areas of
CC protein sequence, the indexer has replaced the residue with an X to
CC represent an illegible residue
XX
SQ Sequence 155 AA;

Query Match 44.3%; Score 47; DB 5; Length 155;
Best Local Similarity 47.1%; Pred. No. 49;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```
Qy          4 LPAVVGLSPGEQEYHRG 20
              ||| :| :| :| : ||
Db         80 LPAFIGANPYQQAFERG 96
```

RESULT 15
AAB69298
ID AAB69298 standard; protein; 192 AA.
XX
AC AAB69298;
XX
DT 12-SEP-2003 (revised)
DT 20-APR-2001 (first entry)
XX
DE HIV-1 non-subtype B clone 96ZM751-3 vif protein.
XX
KW HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW vif; vpr; tat; rev; nef; vaccine.
XX
OS Human immunodeficiency virus 1.
XX
PN WO200026416-A1.
XX
PD 11-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US024837.
XX
PR 02-NOV-1998; 98US-00184418.
XX
PA (UABR-) UAB RES FOUND.
XX
PI Hahn BH, Shaw GM, Gao F;
XX
DR WPI; 2000-365651/31.
XX
PT Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT type 1 useful for detecting and treating AIDS comprises a specific
PT nucleotide sequence.
XX
PS Claim 41; Fig 16; 131pp; English.
XX
CC The present in invention provides the protein and coding sequences for a
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC isolates. The sequences shown include the near full-length coding
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC rev and nef proteins. These can be used to detect the presence of HIV-1
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC These antibodies can be used in vaccines to prevent and treat HIV
CC infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ Sequence 192 AA;

Query Match 44.3%; Score 47; DB 3; Length 192;
Best Local Similarity 66.7%; Pred. No. 62;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
Qy          9 GLSPGEQEYHRG 20
              || |||:|:| |
Db         71 GLHPGEREWHLG 82
```

Search completed: May 7, 2004, 12:33:43
Job time : 56.8 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:27:20 ; Search time 8.96 Seconds
(without alignments)
150.299 Million cell updates/sec

Title: US-09-786-214A-12
Perfect score: 72
Sequence: 1 LPAVVGLSPGEQEY 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	61.1	381	2	F75270	cytochrome P450 -
2	43	59.7	211	2	A64490	precorrin-6Y methy
3	43	59.7	623	2	T40991	probable lysophosp
4	43	59.7	821	2	C84304	DNA helicase [impo
5	41	56.9	156	2	F87551	conserved hypothet
6	41	56.9	326	2	T45226	probable N5,N10-me
7	41	56.9	508	2	E70764	probable cobI prot
8	40	55.6	227	2	B90400	hypothetical prote
9	40	55.6	243	2	I54459	MHC H-2K1-k - mous
10	40	55.6	428	2	AG1304	uracil permease ho
11	40	55.6	428	2	AG1676	uracil permease ho
12	40	55.6	540	2	A75250	carboxylesterase,
13	40	55.6	661	2	G84511	hypothetical prote
14	40	55.6	673	2	T50281	probable lysophosp
15	39	54.2	43	2	S21065	Ig kappa chain V r
16	39	54.2	96	2	S45441	Ig kappa chain V r
17	39	54.2	103	2	S19975	Ig kappa chain V r
18	39	54.2	106	2	PS0070	Ig kappa chain V r
19	39	54.2	106	2	PC4282	Ig kappa chain (an
20	39	54.2	107	2	S57444	Ig kappa chain V-J
21	39	54.2	108	2	C30502	Ig kappa chain V r
22	39	54.2	108	2	S33988	Ig kappa chain V r
23	39	54.2	108	2	G44151	Ig kappa chain V r
24	39	54.2	111	2	S23628	Ig kappa chain V r
25	39	54.2	114	2	S54905	Ig kappa chain V r
26	39	54.2	115	1	K3HUVG	Ig kappa chain pre
27	39	54.2	115	1	KVMSL7	Ig kappa chain pre
28	39	54.2	115	2	S11697	Ig kappa chain pre
29	39	54.2	116	2	B25521	Ig kappa chain pre
30	39	54.2	119	2	S41816	Ig kappa chain V r
31	39	54.2	125	2	S40344	Ig kappa chain V-J
32	39	54.2	128	2	PN0445	Ig kappa chain pre
33	39	54.2	128	2	S40379	Ig kappa chain V-J
34	39	54.2	128	2	A56701	Ig kappa chain V r
35	39	54.2	129	2	S29627	Ig kappa chain V r
36	39	54.2	129	2	S40363	Ig kappa chain - h
37	39	54.2	132	2	S05268	Ig kappa chain pre
38	39	54.2	144	2	PL0106	Ig kappa chain pre
39	39	54.2	144	2	B30502	Ig heavy chain V r
40	39	54.2	215	2	A23746	Ig kappa chain V-I
41	39	54.2	277	2	F82130	formyltetrahydrofo
42	39	54.2	278	2	E64131	formyltetrahydrofo
43	39	54.2	319	2	AD0941	probable regulator
44	39	54.2	563	2	JQ0623	nerve growth facto
45	39	54.2	1367	2	S74285	BUD3 protein - yea

ALIGNMENTS

RESULT 1
F75270
cytochrome P450 - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: F75270
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; , M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: F75270
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-381 <WHI>
A;Cross-references: GB:AE002076; GB:AE000513; NID:g6460285; PIDN:AAF12016.1; PID:g64602
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2473
A;Map position: 1

Query Match 61.1%; Score 44; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPAVVGLSP 9
|||||||||
Db 52 LPAVVGLSP 60

RESULT 2
A64490
precorrin-6Y methylase homolog - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 18-Aug-2003
C;Accession: A64490
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, ; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A. rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschi
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: A64490
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-211 <BUL>
A;Cross-references: GB:U67593; GB:L77117; NID:g2826427; PIDN:AAB99541.1; PID:g1592152;
C;Genetics:
A;Map position: FOR1500322-1500957
C;Superfamily: precorrin-6Y methylase CbiE

```
  Query Match             59.7%;  Score 43;  DB 2;  Length 211;
  Best Local Similarity   54.5%;  Pred. No. 9.3;
  Matches    6;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 VVGLSPGEQEY 14
              :||: ||::||
Db          4 IVGIGPGDREY 14


RESULT 3
T40991
probable lysophospholipase precursor - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C;Accession: T40991
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, March 1999
A;Reference number: Z21962
A;Accession: T40991
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-623 <LYN>
A;Cross-references: EMBL:AL049559; PIDN:CAB40176.1; GSPDB:GN00068; SPDB:SPCC1450.09c
A;Experimental source: strain 972h-; cosmid c1450
C;Genetics:
A;Gene: SPDB:SPCC1450.09c
A;Map position: 3
C;Superfamily: yeast lysophospholipase

  Query Match             59.7%;  Score 43;  DB 2;  Length 623;
  Best Local Similarity   69.2%;  Pred. No. 29;
  Matches    9;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 PAVVGLSPGEQEY 14
              ||  ||| ||||:
Db         76 PASDGLSTGEQEF 88


RESULT 4
C84304
DNA helicase [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: C84304
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
 Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: C84304
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-821 <STO>
A;Cross-references: GB:AE004437; NID:g10580995; PIDN:AAG19799.1; GSPDB:GN00138
C;Genetics:
A;Gene: hel

  Query Match             59.7%;  Score 43;  DB 2;  Length 821;
  Best Local Similarity   90.0%;  Pred. No. 38;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AVVGLSPGEQ 12
              |||||||  ||
Db        326 AVVGLSPAEQ 335


RESULT 5
F87551
conserved hypothetical protein CC2439 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: F87551
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolo
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: F87551
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-156 <STO>
A;Cross-references: GB:AE005673; NID:g13423984; PIDN:AAK24410.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC2439
C;Superfamily: Haemophilus influenzae conserved hypothetical protein HI0305

  Query Match             56.9%;  Score 41;  DB 2;  Length 156;
  Best Local Similarity   80.0%;  Pred. No. 15;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AVVGLSPGEQ 12
              |||||  |||:
Db         18 AVVGLDPGEK 27


RESULT 6
T45226
probable N5,N10-methylene-tetrahydromethanopterin reductase (F420-dependent) [imported]
C;Species: Methanolobus tindarius
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C;Accession: T45226
R;Westenberg, D.J.; Braune, A.; Ruppert, C.; Mueller, V.; Herzberg, C.; Gottschalk, G.;
submitted to the EMBL Data Library, September 1998
A;Description: The F420H2-dehydrogenase from Methanolobus tindarius: Cloning of the ffd
A;Reference number: Z22947
A;Accession: T45226
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-326 <WES>
A;Cross-references: EMBL:AJ011519; PIDN:CAB56639.1
A;Experimental source: DSM 2278
C;Genetics:
A;Gene: ffdA

  Query Match             56.9%;  Score 41;  DB 2;  Length 326;
  Best Local Similarity   70.0%;  Pred. No. 32;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AVVGLSPGEQ 12
              |::||  ||||
Db         88 AILGLGPGEQ 97


RESULT 7
E70764
probable cobI protein - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: E70764
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: E70764
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-508 <COL>
A;Cross-references: GB:Z73966; GB:AL123456; NID:g3261577; PIDN:CAA98214.1; PID:e246996;
A;Experimental source: strain H37Rv
```

```
C;Genetics:
A;Gene: cobI

  Query Match             56.9%;  Score 41;  DB 2;  Length 508;
  Best Local Similarity   58.3%;  Pred. No. 51;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 AVVGLSPGEQEY 14
              ||||| ||: ::
Db        250 AVVGLGPGDSDW 261


RESULT 8
B90400
hypothetical protein cbiE [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 18-Aug-2003
C;Accession: B90400
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: B90400
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-227 <KUR>
A;Cross-references: GB:AE006641; NID:g13815599; PIDN:AAK42457.1; GSPDB:GN00155
C;Genetics:
A;Gene: cbiE
C;Superfamily: precorrin-6Y methylase CbiE

  Query Match             55.6%;  Score 40;  DB 2;  Length 227;
  Best Local Similarity   54.5%;  Pred. No. 32;
  Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 VVGLSPGEQEY 14
              :||: ||: ||
Db         10 IVGVGPGDPEY 20


RESULT 9
I54459
MHC H-2K1-k - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C;Accession: I54459
R;Watts, S.; Davis, A.C.; Goodenow, R.S.
Immunogenetics 29, 355-357, 1989
A;Title: Sequence analysis of the C3H H-2K1-k gene: Relationship to the H-2 loci.
A;Reference number: I54459; MUID:89233303; PMID:2714856
A;Accession: I54459
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-243 <RES>
A;Cross-references: GB:M27134; NID:g199435; PIDN:AAA39610.1; PID:g387456
C;Genetics:
A;Introns: 22/1; 112/1

  Query Match             55.6%;  Score 40;  DB 2;  Length 243;
  Best Local Similarity   77.8%;  Pred. No. 35;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 VGLSPGEQE 13
              :||||||:|
Db        216 LGLSPGEEE 224


RESULT 10
AG1304
uracil permease homolog pyrP [imported] - Listeria monocytogenes (strain EGD-e)
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AG1304
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AG1304
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-428 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC99917.1; PID:g16411293; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: pyrP
C;Superfamily: uracil transport protein uraA

  Query Match             55.6%;  Score 40;  DB 2;  Length 428;
  Best Local Similarity   70.0%;  Pred. No. 63;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPG 10
              :|:| |||||
Db         42 VPSVTGLSPG 51


RESULT 11
AG1676
uracil permease homolog pyrP [imported] - Listeria innocua (strain Clip11262)
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AG1676
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AG1676
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-428 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC97183.1; PID:g16414454; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: pyrP
C;Superfamily: uracil transport protein uraA

  Query Match             55.6%;  Score 40;  DB 2;  Length 428;
  Best Local Similarity   70.0%;  Pred. No. 63;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPG 10
              :|:| |||||
Db         42 VPSVTGLSPG 51


RESULT 12
A75250
carboxylesterase, type B - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: A75250
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
```

A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: A75250
A;Molecule type: DNA
A;Residues: 1-540 <WHI>
A;Cross-references: GB:AE002092; GB:AE000513; NID:g6460455; PIDN:AAF12163.1; PID:g646045
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2626
A;Map position: 1
C;Superfamily: cholinesterase; cholinesterase homology

Query Match 55.6%; Score 40; DB 2; Length 540;
Best Local Similarity 70.0%; Pred. No. 80;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGE 11
| |:||:|||
Db 512 PQVLGLAPGE 521

RESULT 13
G84511
hypothetical protein At2g13900 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: G84511
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: G84511
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-661 <STO>
A;Cross-references: GB:AE002093; NID:g6598598; PIDN:AAF18650.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g13900
A;Map position: 2

Query Match 55.6%; Score 40; DB 2; Length 661;
Best Local Similarity 53.8%; Pred. No. 99;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQEY 14
| : : ||||||
Db 437 PLTLKIVPGEQEY 449

RESULT 14
T50281
probable lysophospholipase (EC 3.1.1.5) precursor SPAC977.09c [similarity] - fission yea
C;Species: Schizosaccharomyces pombe
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 19-Jan-2001
C;Accession: T50281; T42738
R;Zimmermann, W.; Wambutt, R.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, January 2000
A;Reference number: Z25053
A;Accession: T50281
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-673 <ZIM>
A;Cross-references: EMBL:AL137130; NID:g6742151; PIDN:CAB69631.1; PID:g6742159; GSPDB:GN
A;Experimental source: strain 972h(-); cosmid c977
R;Yoshioka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.
DNA Res. 4, 363-369, 1997
A;Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
A;Reference number: Z17323; MUID:98162722; PMID:9501991
A;Accession: T42738
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 427-673 <YOS>
A;Cross-references: EMBL:D89183; NID:g1749573; PIDN:BAA13845.1; PID:g1749574
A;Experimental source: strain PR745
C;Genetics:
A;Gene: SPDB:SPAC977.09c
A;Map position: 1
A;Introns: 651/3
C;Function:
A;Description: catalyzes the hydolysis of 2-lysophosphatidylcholine to glycerophosphoch
C;Superfamily: yeast lysophospholipase
C;Keywords: carboxylic ester hydrolase

Query Match 55.6%; Score 40; DB 2; Length 673;
Best Local Similarity 61.5%; Pred. No. 1e+02;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQEY 14
|| ||: ||| |
Db 83 PASEGLNEGEQSY 95

RESULT 15
S21065
Ig kappa chain V region (anti-RH(D)) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 22-Nov-1993 #sequence_revision 24-May-1996 #text_change 09-May-1997
C;Accession: S21065
R;Dlouha, A.; Lecroisey, A.; Henschen, A.; Rouger, P.; Keil, B.
Protein Seq. Data Anal. 4, 317-318, 1991
A;Title: Subgroup assignement of a human monoclonal anti-Rh(D) antibody.
A;Reference number: S21065; MUID:92253544; PMID:1812483
A;Accession: S21065
A;Molecule type: protein
A;Residues: 1-43 <DLO>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin

Query Match 54.2%; Score 39; DB 2; Length 43;
Best Local Similarity 63.6%; Pred. No. 8.4;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQ 12
|| : |||||:
Db 8 PATLSLSPGER 18

Search completed: May 7, 2004, 12:39:06
Job time : 9.12667 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:20:30 ; Search time 5.04 Seconds
(without alignments)
144.639 Million cell updates/sec

Title: US-09-786-214A-12
Perfect score: 72
Sequence: 1 LPAVVGLSPGEQEY 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	43	59.7	211	1	COBL_METJA	Q58917	methanococc
2	43	59.7	633	1	PLB5_SCHPO	Q9y7n6	schizosacch
3	41	56.9	156	1	RUVX_CAUCR	Q9a5k8	caulobacter
4	41	56.9	326	1	MER_METTI	Q9uxp0	methanolobu
5	41	56.9	508	1	COBI_MYCTU	Q10677	mycobacteri
6	40	55.6	673	1	PLB4_SCHPO	Q9p327	schizosacch
7	39	54.2	115	1	KV3I_HUMAN	P04433	homo sapien
8	39	54.2	115	1	KV5I_MOUSE	P01642	mus musculu
9	39	54.2	278	1	PURU_HAEIN	Q03432	haemophilus
10	39	54.2	516	1	C4AD_DROME	Q9v4t3	drosophila
11	39	54.2	597	1	NR41_RAT	P22829	rattus norv
12	39	54.2	1402	1	N160_MOUSE	Q9z0w3	mus musculu
13	39	54.2	1636	1	BUD3_YEAST	P25558	saccharomyc
14	38	52.8	390	1	COBL_MYCTU	Q10671	mycobacteri
15	38	52.8	429	1	RS1_LEULA	P50889	leuconostoc
16	38	52.8	507	1	CATA_PICAN	P30263	pichia angu
17	38	52.8	699	1	EFG_HAEIN	P43925	haemophilus
18	38	52.8	700	1	EFG_PASMU	P57938	pasteurella
19	37.5	52.1	827	1	M4K1_MOUSE	P70218	mus musculu
20	37	51.4	280	1	PMXA_MOUSE	Q62066	mus musculu
21	37	51.4	281	1	PMXA_RAT	Q62782	rattus norv
22	37	51.4	446	1	COBJ_ARCFU	O29534	a cobalamin
23	37	51.4	446	1	ENO1_MAIZE	P26301	zea mays (m
24	37	51.4	557	1	PUR6_VIGAC	P55195	vigna aconi
25	37	51.4	607	1	GLMS_CLOTE	Q890u2	c glucosami
26	37	51.4	637	1	MUTL_CAUCR	Q9rp66	caulobacter
27	37	51.4	658	1	VG18_BPT4	P13332	bacteriopha
28	37	51.4	753	1	CKAA_BACUF	O32321	bacillus th
29	37	51.4	813	1	CADM_MOUSE	Q9wtp5	mus musculu
30	37	51.4	985	1	4ET_HUMAN	Q9nra8	homo sapien
31	37	51.4	992	1	EVC_HUMAN	P57679	homo sapien
32	37	51.4	2269	1	RRPL_SV41	P35341	simian viru
33	36.5	50.7	374	1	RGSK_BOVIN	P79348	bos taurus
34	36.5	50.7	1121	1	BMS1_SCHPO	O94653	schizosacch
35	36	50.0	129	1	KV3H_HUMAN	P04207	homo sapien
36	36	50.0	158	1	RSD_ECOLI	P31690	escherichia
37	36	50.0	162	1	RSD_SALTY	Q9l9i6	salmonella
38	36	50.0	243	1	SUMT_SYNP7	P42451	synechococc
39	36	50.0	279	1	TYSY_CAUCR	Q9a6h0	caulobacter
40	36	50.0	280	1	PURU_ECOL6	P38480	escherichia
41	36	50.0	280	1	PURU_ECOLI	P37051	escherichia
42	36	50.0	317	1	OAD1_HUMAN	Q8nge0	homo sapien
43	36	50.0	342	1	HUPK_AZOVI	P31878	azotobacter
44	36	50.0	348	1	HOXV_AZOVI	P40597	azotobacter
45	36	50.0	359	1	ALF2_PEA	P46257	pisum sativ

ALIGNMENTS

RESULT 1
COBL_METJA
ID COBL_METJA STANDARD; PRT; 211 AA.
AC Q58917;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable precorrin-6Y C5,15-methyltransferase [decarboxylating]
DE (EC 2.1.1.132) (Precorrin-6 methyltransferase) (Precorrin-6Y
DE methylase).
GN COBL OR MJ1522.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: CATALYZES THE METHYLATION OF BOTH C-5 AND C-15 IN
CC PRECORRIN-6Y TO FORM PRECORRIN-8X (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 2 S-adenosyl-L-methionine + precorrin-6Y = 2
CC S-adenosyl-L-homocysteine + precorrin-8X + CO(2).
CC -!- PATHWAY: Cobalamin biosynthesis.
CC -!- SIMILARITY: TO S.TYPHIMURIUM CBIE; ALSO, LOW, TO OTHER
CC METHYLASES INVOLVED IN COBALAMIN BIOSYNTHESIS.
CC ---

CC ---
DR EMBL; U67593; AAB99541.1; -.
DR PIR; A64490; A64490.
DR TIGR; MJ1522; -.
DR InterPro; IPR000878; Cor/por_Metransf.
DR Pfam; PF00590; TP_methylase; 1.
KW Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW Methyltransferase; Complete proteome.
SQ SEQUENCE 211 AA; 23805 MW; 279A1A2B14369510 CRC64;

Query Match 59.7%; Score 43; DB 1; Length 211;

```
 Best Local Similarity   54.5%;  Pred. No. 5.8;
 Matches    6;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 VVGLSPGEQEY 14
              :||: ||::||
Db          4 IVGIGPGDREY 14


RESULT 2
PLB5_SCHPO
ID   PLB5_SCHPO     STANDARD;      PRT;   633 AA.
AC   Q9Y7N6;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Putative lysophospholipase C1450.09c precursor (EC 3.1.1.5)
DE   (Phospholipase B).
GN   SPCC1450.09C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: Catalyzes the release of fatty acids from
CC       lysophospholipids (By similarity).
CC   -!- CATALYTIC ACTIVITY: 2-lysophosphatidylcholine + H(2)O =
CC       glycerophosphocholine + a fatty acid anion.
CC   -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC   -!- SIMILARITY: Belongs to the lysophospholipase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL049559; CAB40176.2; -.
DR   GeneDB_SPombe; SPCC1450.09c; -.
DR   InterPro; IPR002642; PLAc.
DR   Pfam; PF01735; PLA2_B; 1.
DR   SMART; SM00022; PLAc; 1.
KW   Hypothetical protein; Lipid degradation; Hydrolase; Glycoprotein;
KW   Signal.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    633       PUTATIVE LYSOPHOSPHOLIPASE C1450.09C.
FT   CARBOHYD    118    118       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    153    153       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    232    232       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    256    256       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    264    264       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    293    293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    331    331       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    360    360       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    367    367       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    400    400       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    403    403       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    474    474       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    508    508       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    513    513       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    537    537       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    564    564       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    586    586       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    603    603       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   633 AA;  68292 MW;  49871B2955893D19 CRC64;

  Query Match             59.7%;  Score 43;  DB 1;  Length 633;
  Best Local Similarity   69.2%;  Pred. No. 17;
  Matches    9;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 PAVVGLSPGEQEY 14
              ||  ||| ||||:
Db         76 PASDGLSTGEQEF 88


RESULT 3
RUVX_CAUCR
ID   RUVX_CAUCR     STANDARD;      PRT;   156 AA.
AC   Q9A5K8;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Putative Holliday junction resolvase (EC 3.1.-.-).
GN   CC2439.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC   Caulobacteraceae; Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC   -!- FUNCTION: Could be a nuclease that resolves Holliday junction
CC       intermediates in genetic recombination.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: Belongs to the yqgF HJR family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; AE005913; AAK24410.1; -.
DR   PIR; F87551; F87551.
DR   TIGR; CC2439; -.
DR   HAMAP; MF_00651; -; 1.
DR   InterPro; IPR005227; Cons_hypoth250.
DR   InterPro; IPR006641; YqgFc.
DR   Pfam; PF03652; UPF0081; 1.
DR   SMART; SM00732; YqgFc; 1.
DR   TIGRFAMs; TIGR00250; TIGR00250; 1.
KW   Hydrolase; Nuclease; DNA repair; DNA recombination; Complete proteome.
SQ   SEQUENCE   156 AA;  17142 MW;  21F54D8648396141 CRC64;

  Query Match             56.9%;  Score 41;  DB 1;  Length 156;
  Best Local Similarity   80.0%;  Pred. No. 9.4;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AVVGLSPGEQ 12
              ||||| |||:
Db         18 AVVGLDPGEK 27


RESULT 4
MER_METTI
ID   MER_METTI      STANDARD;      PRT;   326 AA.
AC   Q9UXP0;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Coenzyme F420-dependent N(5),N(10)-methylenetetrahydromethanopterin
DE   reductase (EC 1.5.99.11) (Methylene-H(4)MPT reductase).
GN   MER OR FFDA.
OS   Methanolobus tindarius.
OC   Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC   Methanosarcinaceae; Methanolobus.
OX   NCBI_TaxID=2221;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 2278;
RX   MEDLINE=99132696; PubMed=9933933;
RA   Westenberg D.J., Braune A., Ruppert C., Mueller V., Herzberg C.,
RA   Gottschalk G., Blaut M.;
RT   "The F420H2-dehydrogenase from Methanolobus tindarius: cloning of the
RT   ffd operon and expression of the genes in Escherichia coli.";
RL   FEMS Microbiol. Lett. 170:389-398(1999).
CC   -!- FUNCTION: Catalyzes the reversible reduction of methylene-H(4)MPT
CC       to methyl-H(4)MPT (By similarity).
CC   -!- CATALYTIC ACTIVITY: N(5),N(10)-methylenetetrahydromethanopterin +
CC       reduced coenzyme F420 = 5-methyl-5,6,7,8-tetrahydromethanopterin +
CC       coenzyme F420.
CC   -!- PATHWAY: Methanogenesis from carbon dioxide; fifth step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the mer family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ011519; CAB56639.1; -.
DR   PIR; T45226; T45226.
DR   HAMAP; MF_01091; -; 1.
DR   InterPro; IPR002103; Bac_luciferase.
DR   Pfam; PF00296; bac_luciferase; 1.
KW   Methanogenesis; One-carbon metabolism; Oxidoreductase.
SQ   SEQUENCE   326 AA;  34043 MW;  16F3AB9733A45D82 CRC64;

  Query Match             56.9%;  Score 41;  DB 1;  Length 326;
  Best Local Similarity   70.0%;  Pred. No. 19;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AVVGLSPGEQ 12
              |::|| ||||
Db         88 AILGLGPGEQ 97


RESULT 5
COBI_MYCTU
ID   COBI_MYCTU     STANDARD;      PRT;   508 AA.
AC   Q10677;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Cobalamin biosynthesis protein COBIJ [Includes: Precorrin-2 C20-
DE   methyltransferase (EC 2.1.1.130) (S-adenosyl-L-methionine--precorrin-2
DE   methyltransferase) (SP2MT); Precorrin-3 methylase (EC 2.1.1.-)].
GN   COBIJ OR COBI OR RV2066 OR MT2126 OR MTCY49.05 OR MB2092.
OS   Mycobacterium tuberculosis, and
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773, 1765;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX   MEDLINE=22206494; PubMed=12218036;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT   "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   J. Bacteriol. 184:5479-5490(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.bovis; STRAIN=AF2122/97;
RX   MEDLINE=22709107; PubMed=12788972;
RA   Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA   Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA   Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA   Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT   "The complete genome sequence of Mycobacterium bovis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC   -!- FUNCTION: METHYLATES PRECORRIN-2 AT THE C-20 POSITION TO PRODUCE
CC       PRECORRIN-3A (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-2 = S-
CC       adenosyl-L-homocysteine + precorrin-3A.
CC   -!- PATHWAY: Cobalamin biosynthesis.
CC   -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS SUMT, CYSG, CBIF/COBM
CC       AND CBIL/COBI.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ----------------------------------------------------------------------
DR   EMBL; Z73966; CAA98214.1; -.
DR   EMBL; AE007063; AAK46406.1; -.
DR   EMBL; BX248341; CAD96945.1; -.
DR   PIR; E70764; E70764.
DR   TIGR; MT2126; -.
DR   TubercuList; Rv2066; -.
DR   InterPro; IPR006364; CobI_CbiL.
DR   InterPro; IPR006363; CobJ.
DR   InterPro; IPR000878; Cor/por_Metransf.
DR   InterPro; IPR003043; Uropor_Metransf.
DR   Pfam; PF00590; TP_methylase; 2.
DR   TIGRFAMs; TIGR01467; cobI_cbiL; 1.
DR   TIGRFAMs; TIGR01466; cobJ_cbiH; 1.
DR   PROSITE; PS00839; SUMT_1; 1.
DR   PROSITE; PS00840; SUMT_2; 1.
KW   Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW   Methyltransferase; Multifunctional enzyme; Complete proteome.
FT   DOMAIN        1    243       PRECORRIN-2 C20-METHYLTRANSFERASE.
FT   DOMAIN      244    508       PRECORRIN-3 METHYLASE.
SQ   SEQUENCE   508 AA;  53910 MW;  95AC066F022C4DC1 CRC64;

  Query Match             56.9%;  Score 41;  DB 1;  Length 508;
  Best Local Similarity   58.3%;  Pred. No. 30;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 AVVGLSPGEQEY 14
              ||||| ||: ::
Db        250 AVVGLGPGDSDW 261


RESULT 6
PLB4_SCHPO
ID   PLB4_SCHPO     STANDARD;      PRT;   673 AA.
AC   Q9P327;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Putative lysophospholipase C977.09c precursor (EC 3.1.1.5)
DE   (Phospholipase B).
GN   SPAC977.09C OR SPAC1348.10C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: Catalyzes the release of fatty acids from
CC       lysophospholipids (By similarity).
CC   -!- CATALYTIC ACTIVITY: 2-lysophosphatidylcholine + H(2)O =
CC       glycerophosphocholine + a fatty acid anion.
CC   -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC   -!- SIMILARITY: Belongs to the lysophospholipase family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AL358912; CAB94277.1; -.
DR   EMBL; AL137130; CAB69631.1; -.
DR   PIR; T50281; T50281.
DR   GeneDB_SPombe; SPAC977.09c; -.
DR   InterPro; IPR001179; FKBP_PPIase.
DR   InterPro; IPR002642; PLAc.
DR   Pfam; PF01735; PLA2_B; 1.
DR   SMART; SM00022; PLAc; 1.
KW   Hypothetical protein; Lipid degradation; Hydrolase; Glycoprotein;
KW   Signal.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    673       PUTATIVE LYSOPHOSPHOLIPASE C977.09C.
FT   CARBOHYD     72     72       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    191    191       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    194    194       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    301    301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    374    374       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    404    404       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    409    409       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    481    481       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    516    516       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    545    545       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    574    574       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   673 AA;  74595 MW;  B39A773E76CD694B CRC64;

  Query Match             55.6%;  Score 40;  DB 1;  Length 673;
  Best Local Similarity   61.5%;  Pred. No. 58;
  Matches    8;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 PAVVGLSPGEQEY 14
              ||  ||: ||| |
Db         83 PASEGLNEGEQSY 95


RESULT 7
KV3I_HUMAN
ID   KV3I_HUMAN     STANDARD;      PRT;   115 AA.
AC   P04433;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Ig kappa chain V-III region VG precursor (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85087932; PubMed=6440122;
RA   Pech M., Zachau H.G.;
RT   "Immunoglobulin genes of different subgroups are interdigitated
RT   within the VK locus.";
RL   Nucleic Acids Res. 12:9229-9236(1984).
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X01668; -; NOT_ANNOTATED_CDS.
DR   PIR; A01900; K3HUVG.
DR   HSSP; P80362; 1WTL.
DR   GO; GO:0005576; C:extracellular; NAS.
DR   GO; GO:0003823; F:antigen binding; NAS.
DR   GO; GO:0006955; P:immune response; NAS.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Immunoglobulin V region; Signal.
FT   SIGNAL        1     20
FT   CHAIN        21   >115       IG KAPPA CHAIN V-III REGION VG.
FT   DOMAIN       21     43       FRAMEWORK-1.
FT   DOMAIN       44     54       COMPLEMENTARITY-DETERMINING-1.
FT   DOMAIN       55     69       FRAMEWORK-2.
FT   DOMAIN       70     76       COMPLEMENTARITY-DETERMINING-2.
FT   DOMAIN       77    108       FRAMEWORK-3.
FT   DOMAIN      109    115       COMPLEMENTARITY-DETERMINING-3.
FT   DISULFID     43    108       BY SIMILARITY.
FT   NON_TER     115    115
SQ   SEQUENCE   115 AA;  12575 MW;  2DE47CDA3A17D555 CRC64;

  Query Match             54.2%;  Score 39;  DB 1;  Length 115;
  Best Local Similarity   63.6%;  Pred. No. 15;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 PAVVGLSPGEQ 12
              || : ||||:
Db         28 PATLSLSPGER 38


RESULT 8
KV5I_MOUSE
ID   KV5I_MOUSE     STANDARD;      PRT;   115 AA.
AC   P01642;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Ig kappa chain V-V region L7 precursor (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=81220975; PubMed=6264318;
RA   Pech M., Hochtl J., Schnell H., Zachau H.G.;
RT   "Differences between germ-line and rearranged immunoglobulin V kappa
RT   coding sequences suggest a localized mutation mechanism.";
RL   Nature 291:668-670(1981).
CC   -!- MISCELLANEOUS: THERE APPEAR TO BE TWO POSSIBLE SPLICE JUNCTIONS AT
CC       THE 3' END OF THE INTRON. THE ALTERNATE WOULD CODE FOR A PROTEIN
CC       LACKING RESIDUES 17-19.
DR   PIR; A01925; KVMSL7.
DR   PDB; 1J1O; 18-FEB-03.
DR   PDB; 1J1P; 18-FEB-03.
DR   PDB; 1J1X; 18-FEB-03.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Immunoglobulin V region; Signal; 3D-structure.
FT   SIGNAL        1     20
FT   CHAIN        21   >115       IG KAPPA CHAIN V-V REGION L7.
FT   DOMAIN       21     43       FRAMEWORK-1.
FT   DOMAIN       44     54       COMPLEMENTARITY-DETERMINING-1.
FT   DOMAIN       55     69       FRAMEWORK-2.
FT   DOMAIN       70     76       COMPLEMENTARITY-DETERMINING-2.
FT   DOMAIN       77    108       FRAMEWORK-3.
FT   DOMAIN      109   >115       COMPLEMENTARITY-DETERMINING-3.
FT   DISULFID     43    108       BY SIMILARITY.
FT   NON_TER     115    115
SQ   SEQUENCE   115 AA;  12615 MW;  C17BEC758C577E00 CRC64;

  Query Match             54.2%;  Score 39;  DB 1;  Length 115;
  Best Local Similarity   54.5%;  Pred. No. 15;
  Matches    6;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 PAVVGLSPGEQ 12
              ||:: :||||:
Db         28 PAILSVSPGER 38


RESULT 9
PURU_HAEIN
ID   PURU_HAEIN     STANDARD;      PRT;   278 AA.
AC   Q03432;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-FH(4)
DE   hydrolase).
GN   PURU OR HI1588.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
RN   [2]
RP   SEQUENCE OF 64-278 FROM N.A.
RC   STRAIN=RM 7004 / Serotype B;
RX   MEDLINE=93328119; PubMed=8335255;
RA   Maskell D.J.;
RT   "Cloning and sequencing of the Haemophilus influenzae aroA gene.";
RL   Gene 129:155-156(1993).
CC   -!- FUNCTION: PRODUCES FORMATE FROM FORMYL-TETRAHYDROFOLATE. PROVIDES
CC       THE MAJOR SOURCE OF FORMATE FOR THE PURT-DEPENDENT SYNTHESIS OF
CC       5'-PHOSPHORIBOSYL-N-FORMYLGLYCINAMIDE (FGAR) DURING AEROBIC
CC       GROWTH. HAS A ROLE IN REGULATING THE ONE-CARBON POOL
CC       (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + H(2)O = formate +
CC       tetrahydrofolate.
CC   -!- ENZYME REGULATION: Activated by methionine, inhibited by glycine
CC       (By similarity).
CC   -!- PATHWAY: De novo purine biosynthesis.
CC   -!- SUBUNIT: Homohexamer (By similarity).
CC   -!- SIMILARITY: SOME, TO GAR TRANSFORMYLASE (PURN).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U32833; AAC23236.1; -.
DR   EMBL; L04686; AAA24942.1; -.
DR   PIR; E64131; E64131.
DR   HSSP; P08179; 1GRC.
DR   TIGR; HI1588; -.
DR   InterPro; IPR002912; ACT.
DR   InterPro; IPR002376; formyl_transf.
DR   InterPro; IPR004810; PurU.
DR   Pfam; PF01842; ACT; 1.
DR   Pfam; PF00551; formyl_transf; 1.
DR   PRINTS; PR01575; FFH4HYDRLASE.
DR   TIGRFAMs; TIGR00655; PurU; 1.
KW   Purine biosynthesis; Hydrolase; One-carbon metabolism;
KW   Complete proteome.
FT   ACT_SITE    223    223       BY SIMILARITY.
FT   CONFLICT    115    117       VIG -> RNR (IN REF. 2).
FT   CONFLICT    138    140       HEN -> PK (IN REF. 2).
FT   CONFLICT    205    205       K -> E (IN REF. 2).
SQ   SEQUENCE   278 AA;  32173 MW;  7F375AB3C422EC4B CRC64;

  Query Match             54.2%;  Score 39;  DB 1;  Length 278;
  Best Local Similarity   50.0%;  Pred. No. 36;
  Matches    7;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQEY 14
              ||| :|  | :| |
Db        191 LPAFIGAKPYQQAY 204


RESULT 10
C4AD_DROME
ID   C4AD_DROME     STANDARD;      PRT;   516 AA.
AC   Q9V4T3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable cytochrome P450 4ad1 (EC 1.14.-.-) (CYPIVAD1).
GN   CYP4AD1 OR CG2110.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley; TISSUE=Embryo;
RX   MEDLINE=22426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC   -!- FUNCTION: May be involved in the metabolism of insect hormones and
CC       in the breakdown of synthetic insecticides (By similarity).
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum
CC       (Potential).
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE003837; AAF59092.1; -.
DR   EMBL; AY061058; AAL28606.1; -.
DR   HSSP; P14779; 1JPZ.
DR   FlyBase; FBgn0033292; Cyp4ad1.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW   Endoplasmic reticulum; Hypothetical protein.
FT   METAL       445    445       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   516 AA;  58870 MW;  648EA22492AF58C7 CRC64;

  Query Match             54.2%;  Score 39;  DB 1;  Length 516;
  Best Local Similarity   57.1%;  Pred. No. 65;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQEY 14
              |||| || ||  ::
Db        469 LPAVDGLPPGINDH 482


RESULT 11
NR41_RAT
ID   NR41_RAT       STANDARD;      PRT;   597 AA.
AC   P22829;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
```

```
DE   Orphan nuclear receptor HMR (Nerve growth factor induced protein I-B)
DE   (NGFI-B) (NUR77).
GN   NR4A1 OR HMR OR NGFIB.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90166506; PubMed=3272167;
RA   Milbrandt J.;
RT   "Nerve growth factor induces a gene homologous to the glucocorticoid
RT   receptor gene.";
RL   Neuron 1:183-188(1988).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=93361012; PubMed=8395013;
RA   Wilson T.E., Fahrner T.J., Milbrandt J.;
RT   "The orphan receptors NGFI-B and steroidogenic factor 1 establish
RT   monomer binding as a third paradigm of nuclear receptor-DNA
RT   interaction.";
RL   Mol. Cell. Biol. 13:5794-5804(1993).
RN   [3]
RP   DNA BINDING MOTIFS.
RX   MEDLINE=92229411; PubMed=1314418;
RA   Wilson T.E., Paulsen R.E., Padgett K.A., Milbrandt J.;
RT   "Participation of non-zinc finger residues in DNA binding by two
RT   nuclear orphan receptors.";
RL   Science 256:107-110(1992).
RN   [4]
RP   PHOSPHORYLATION.
RX   MEDLINE=94043340; PubMed=8227042;
RA   Hirata Y., Kiuchi K., Chen H.-C., Milbrandt J., Guroff G.;
RT   "The phosphorylation and DNA binding of the DNA-binding domain of the
RT   orphan nuclear receptor NGFI-B.";
RL   J. Biol. Chem. 268:24808-24812(1993).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 264-351 IN COMPLEX WITH NBRE,
RP   AND METAL-BINDING.
RX   MEDLINE=99260743; PubMed=10331876;
RA   Meinke G., Sigler P.B.;
RT   "DNA-binding mechanism of the monomeric orphan nuclear receptor
RT   NGFI-B.";
RL   Nat. Struct. Biol. 6:471-477(1999).
CC   -!- FUNCTION: Probable nuclear receptor. May act concomitantly with
CC       NURR1 in regulating the expression of delayed-early genes during
CC       liver regeneration. Binds the NGFI-B response element (NBRE) 5'-
CC       AAAAGGTCA-3'.
CC   -!- SUBUNIT: Binds DNA as a monomer.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Expressed in lung, brain and superior
CC       cervical ganglia. High levels are seen in the adrenal tissue.
CC   -!- INDUCTION: By nerve growth factor and during liver regeneration.
CC   -!- PTM: Phosphorylation of Ser-350 results in decrease in NBRE
CC       binding while phosphorylation of Ser-340 has little effect on it.
CC   -!- SIMILARITY: Belongs to the nuclear hormone receptor family. NR4
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17254; AAA56770.1; ALT_INIT.
DR   PDB; 1CIT; 26-JUN-00.
DR   TRANSFAC; T00619; -.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR008946; Str_ncl_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Phosphorylation; 3D-structure.
FT   DNA_BIND     266     331       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING      266     286       C4-TYPE.
FT   ZN_FING      302     326       C4-TYPE.
FT   DOMAIN       408     458       LIGAND-BINDING (POTENTIAL).
FT   DOMAIN        80      91       POLY-SER.
FT   DOMAIN       182     186       POLY-PRO.
FT   DOMAIN       582     585       POLY-PRO.
FT   MOD_RES      340     340       PHOSPHORYLATION (BY PKA).
FT   MOD_RES      350     350       PHOSPHORYLATION (BY PKA).
FT   MUTAGEN      340     340       S->A: LOSS OF PHOSPHORYLATION.
FT   MUTAGEN      350     350       S->A: LOSS OF PHOSPHORYLATION.
FT   MUTAGEN      345     345       R->K: DECREASED NBRE BINDING.
FT   MUTAGEN      348     348       L->V: ALMOST COMPLETE LOSS OF NBRE
FT                                  BINDING.
SQ   SEQUENCE   597 AA;  64281 MW;  9CFA987112337E53 CRC64;

  Query Match             54.2%;  Score 39;  DB 1;  Length 597;
  Best Local Similarity   46.2%;  Pred. No. 75;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQE 13
              :|  : ||||:|:
Db        423 IPGFIELSPGDQD 435


RESULT 12
N160_MOUSE
ID   N160_MOUSE     STANDARD;      PRT;  1402 AA.
AC   Q9Z0W3; Q9CZD9;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Nuclear pore complex protein Nup160 (Nucleoporin Nup160) (160 kDa
DE   nucleoporin) (Gene trap locus 1-13) (GTL-13).
GN   NUP160 OR GTL1-13 OR KIAA0197.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SvJ;
RA   Van de Putte T., Cozijnsen M., Dewulf N., Tylzanowski P., Lonnoy O.,
RA   Huylebroeck D.;
RT   "Mus musculus mRNA for gtl-13 (gene trap locus-13), similar to human
RT   KIAA0197 gene (D83781), complete cds.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1151-1402 FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
```

```
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   IDENTIFICATION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX   MEDLINE=21448620; PubMed=11564755;
RA   Belgareh N., Rabut G., Bai S.W., van Overbeek M., Beaudouin J.,
RA   Daigle N., Zatsepina O.V., Pasteau F., Labas V., Fromont-Racine M.,
RA   Ellenberg J., Doye V.;
RT   "An evolutionarily conserved NPC subcomplex, which redistributes in
RT   part to kinetochores in mammalian cells.";
RL   J. Cell Biol. 154:1147-1160(2001).
RN   [4]
RP   IDENTIFICATION, FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX   MEDLINE=21541555; PubMed=11684705;
RA   Vasu S., Shah S., Orjalo A., Park M., Fischer W.H., Forbes D.J.;
RT   "Novel vertebrate nucleoporins Nup133 and Nup160 play a role in mRNA
RT   export.";
RL   J. Cell Biol. 155:339-354(2001).
CC   -!- FUNCTION: Involved in poly(A)+ RNA transport.
CC   -!- SUBUNIT: Forms part of the Nup160 subcomplex in the nuclear pore
CC       which is composed of Nup160, Nup133, Nup107 and Nup96. This
CC       complex plays a role in RNA export and in tethering Nup98 and
CC       Nup153 to the nucleus.
CC   -!- SUBCELLULAR LOCATION: Nuclear pore complex.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC       frameshift in position 1157 and a stop codon in position 1396.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF104415; AAD17922.2; -.
DR   EMBL; AK012715; BAB28429.1; ALT_FRAME.
DR   MGD; MGI:1926227; Nup160.
DR   GO; GO:0005643; C:nuclear pore; IDA.
DR   GO; GO:0005487; F:nucleocytoplasmic transporter activity; IDA.
DR   GO; GO:0006406; P:mRNA-nucleus export; IDA.
KW   Nuclear protein; Transport.
FT   CONFLICT   1156   1156       A -> T (IN REF. 2).
FT   CONFLICT   1314   1314       E -> G (IN REF. 2).
FT   CONFLICT   1368   1368       N -> D (IN REF. 2).
SQ   SEQUENCE   1402 AA;  158230 MW;  3BF5D9F057D28772 CRC64;

  Query Match             54.2%;  Score 39;  DB 1;  Length 1402;
  Best Local Similarity   70.0%;  Pred. No. 1.7e+02;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPG 10
              :|:| |||||
Db        170 IPSVPGLSPG 179


RESULT 13
BUD3_YEAST
ID   BUD3_YEAST     STANDARD;      PRT;  1636 AA.
AC   P25558; P25556; P25557; P87007;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Bud site selection protein BUD3.
GN   BUD3 OR YCL014W/YCL013W/YCL012W OR YCL14W/YCL13W/YCL12W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95247824; PubMed=7730410;
RA   Chant J., Mischke M., Mitchell E., Herskowitz I., Pringle J.R.;
RT   "Role of Bud3p in producing the axial budding pattern of yeast.";
RL   J. Cell Biol. 129:767-778(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=92244356; PubMed=1574125;
RA   Oliver S.G., van der Aart Q.J.M., Agostoni-Carbone M.L., Aigle M.,
RA   Alberghina L., Alexandraki D., Antoine G., Anwar R., Ballesta J.P.G.,
RA   Benit P., Berben G., Bergantino E., Biteau N., Bolle P.-A.,
RA   Bolotin-Fukuhara M., Brown A.J.P., Brown R., Buhler J.-M.,
RA   Carignani G., Chanet R., Contreras R., Crouzet M., Daignan-Fornier B.,
RA   De Haan M., Defoor E., Delgado M.D., Demolder J., Doira C., Dubois E.,
RA   Dujon B., Duesterhoeft A., Erdmann D., Esteban M., Fabre F.,
RA   Fairhead C.A., Faye G., Feldmann H., Fiers W.,
RA   Francingues-Gaillard M.-C., Franco L., Frontali L., Fukuhara H.,
RA   Fuller L.J., Gent M.E., Gigot D., Gilliquet V., Glansdorff N.,
RA   Goffeau A., Grenson M., Grisanti P., Grivell L.A., Haasemann M.,
RA   Hatat D., Hegemann J.H., Herbert C.J., Hilger F., Hohmann S.,
RA   Hollenberg C.P., Huse K., Iborra F., Indge K.J., Isono K., Jackman P.,
RA   Jacq C., Jacquet M., James C.M., Jauniaux J.-C., Jia Y., Jimenez A.,
RA   Kleinhans U., Kreisl P., Lafranchi G., Lewis C., van der Linden C.G.,
RA   Lucchini G., Lutzenkirchen K., Maat C., Mannhaupt G., Manzano M.E.,
RA   Martegani E., Mathieu A., Maurer C.T.C., McConnell D., McKee R.A.,
RA   Messenguy F., Mewes H.-W., Molemans F., Montague M.A., Navas L.,
RA   Newlon C.S., Olson M.V., Pallier C., Panzeri L., Pearson B.M.,
RA   Perea J., Philippsen P., Pierard A., Planta R.J., Plevani P.,
RA   Poetsch B., Pohl F.M., Purnelle B., Ramezani Rad M., Rasmussen S.W.,
RA   Raynal A., Remacha M., Richterich P., Roberts A.B., Rodriguez F.,
RA   Sanz E., Schaaff-Gerstenschlaeger I., Scherens B., Schweitzer B.,
RA   Shu Y., Skala J., Slonimski P.P., Sor F., Soustelle C.,
RA   Spiegelberg R., Stateva L.I., Steensma H.Y., Steiner S., Thierry A.,
RA   Thireos G., Triano L.N., Urrestarazu L.A., Valle G., Vetter I.,
RA   van Vliet-Reedijk J.C., Volckaert G., Vreken P., Warmington J.R.,
RA   von Wettstein D., Wicksteed B.L., Wilson C., Wurst H., Xu G.,
RA   Zimmermann F.K., Sgouros J.G.;
RT   "The complete DNA sequence of yeast chromosome III.";
RL   Nature 357:38-46(1992).
RN   [3]
RP   REVISIONS.
RA   Gromadka R.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   REVISIONS.
RA   Valles G., Volckaerts G.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Co-assembles with BUD4 at bud sites. BUD4 and BUD3 may
CC       cooperate to recognize a spatial landmark (the neck filaments)
CC       during mitosis and they subsequently become a landmark for
CC       establishing the axial budding pattern in G1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17580; AAA86315.1; -.
DR   EMBL; X59720; CAA42346.2; -.
DR   PIR; S74285; S74285.
DR   PIR; S74286; S74286.
DR   GermOnline; 138855; -.
DR   SGD; S0000520; BUD3.
DR   InterPro; IPR000219; RhoGEF.
DR   SMART; SM00325; RhoGEF; 1.
KW   Cell cycle.
```

```
SQ   SEQUENCE   1636 AA;  184717 MW;  9E4E46BA5C3A3F69 CRC64;

  Query Match             54.2%;  Score 39;  DB 1;  Length 1636;
  Best Local Similarity   58.3%;  Pred. No. 2e+02;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 PAVVGLSPGEQE 13
              ||:  ||| :||
Db        112 PAIENLSPSDQE 123


RESULT 14
COBL_MYCTU
ID   COBL_MYCTU     STANDARD;      PRT;   390 AA.
AC   Q10671;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Precorrin-6Y C5,15-methyltransferase [decarboxylating] (EC 2.1.1.132)
DE   (Precorrin-6 methyltransferase) (Precorrin-6Y methylase).
GN   COBL OR RV2072C OR MT2132 OR MTCY49.11C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RX   MEDLINE=22206494; PubMed=12218036;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT   "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   J. Bacteriol. 184:5479-5490(2002).
CC   -!- FUNCTION: CATALYZES THE METHYLATION OF BOTH C-5 AND C-15 IN
CC       PRECORRIN-6Y TO FORM PRECORRIN-8X.
CC   -!- CATALYTIC ACTIVITY: 2 S-adenosyl-L-methionine + precorrin-6Y = 2
CC       S-adenosyl-L-homocysteine + precorrin-8X + CO(2).
CC   -!- PATHWAY: Cobalamin biosynthesis.
CC   -!- SIMILARITY: TO S.TYPHIMURIUM CBIE; ALSO, LOW, TO OTHER
CC       METHYLASES INVOLVED IN COBALAMIN BIOSYNTHESIS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z73966; CAA98225.1; -.
DR   EMBL; AE007063; AAK46412.1; -.
DR   PIR; C70765; C70765.
DR   TIGR; MT2132; -.
DR   TubercuList; Rv2072c; -.
DR   InterPro; IPR006365; CobL.
DR   InterPro; IPR000878; Cor/por_Metransf.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF00590; TP_methylase; 1.
DR   TIGRFAMs; TIGR01468; cobL_cbiET; 1.
KW   Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW   Methyltransferase; Complete proteome.
FT   CONFLICT    205    205       L -> P (IN REF. 2).
FT   CONFLICT    327    327       D -> H (IN REF. 2).
SQ   SEQUENCE   390 AA;  41854 MW;  FB42EFF7562F21F3 CRC64;

  Query Match             52.8%;  Score 38;  DB 1;  Length 390;
  Best Local Similarity   88.9%;  Pred. No. 73;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSP 9
              |||| ||||
Db         55 LPAVQGLSP 63


RESULT 15
RS1_LEULA
ID   RS1_LEULA      STANDARD;      PRT;   429 AA.
AC   P50889; P71450;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   40S ribosomal protein S1.
GN   RPS1.
OS   Leuconostoc lactis.
OC   Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX   NCBI_TaxID=1246;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97186703; PubMed=9034319;
RA   Yamit-Hezi A., Levy Z., Neuman S., Nudel U.;
RT   "A Leuconostoc lactis protein with homology to ribosomal protein S1
RT   shares common epitopes and common DNA binding properties with a
RT   mammalian DNA binding nuclear factor.";
RL   Gene 185:99-103(1997).
RN   [2]
RP   SEQUENCE OF 24-429 FROM N.A.
RX   MEDLINE=95237615; PubMed=7721096;
RA   Eklund E.A., Lee S.W., Skalnik D.G.;
RT   "Cloning of a cDNA encoding a human DNA-binding protein similar to
RT   ribosomal protein S1.";
RL   Gene 155:231-235(1995).
RN   [3]
RP   SEQUENCE OF 78-429 FROM N.A.
RX   MEDLINE=96164600; PubMed=8568274;
RA   Tsuzaka K., Leu A.K., Frank M.B., Movafagh B.F., Koscec M.,
RA   Winkler T.H., Kalden J.R., Reichlin M.;
RT   "Lupus autoantibodies to double-stranded DNA cross-react with
RT   ribosomal protein S1.";
RL   J. Immunol. 156:1668-1675(1996).
CC   -!- FUNCTION: EXHIBITS PREFERENTIAL BINDING TO SINGLE-STRANDED AND
CC       DOUBLE-STRANDED DNA AND A LOW BINDING AFFINITY FOR RNA.
CC   -!- SIMILARITY: Belongs to the S1P family of ribosomal proteins.
CC   -!- SIMILARITY: Contains 4 S1 motif domains.
CC   -!- CAUTION: WAS ORIGINALLY (REF.2 AND REF.3) THOUGHT TO ORIGINATE
CC       FROM HUMAN BUT IS MOST PROBABLY THE RESULT OF A CDNA LIBRARY
CC       CONTAMINATION BY L.LACTIS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U24086; AAB08978.1; -.
DR   EMBL; U05589; AAA77669.1; -.
```

```
DR   EMBL; U27517; AAA97575.1; -.
DR   HSSP; P05055; 1SRO.
DR   InterPro; IPR008994; Nucleic_acid_OB.
DR   InterPro; IPR000110; Ribosomal_S1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF00575; S1; 4.
DR   PRINTS; PR00681; RIBOSOMALS1.
DR   SMART; SM00316; S1; 4.
DR   PROSITE; PS50126; S1; 4.
KW   Ribosomal protein; Repeat; RNA-binding.
FT   DOMAIN       55    128       S1 MOTIF 1.
FT   DOMAIN      144    211       S1 MOTIF 2.
FT   DOMAIN      231    299       S1 MOTIF 3.
FT   DOMAIN      316    385       S1 MOTIF 4.
FT   CONFLICT     24     24       S -> G (IN REF. 2).
FT   CONFLICT    122    122       A -> S (IN REF. 3).
FT   CONFLICT    217    217       L -> R (IN REF. 2 AND 3).
SQ   SEQUENCE   429 AA;  46386 MW;  92AC82605F39DDFC CRC64;

  Query Match             52.8%;  Score 38;  DB 1;  Length 429;
  Best Local Similarity   80.0%;  Pred. No. 80;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AVVGLSPGEQ 12
              |||||| ||:
Db         71 AVVGLSTGEE 80


Search completed: May  7, 2004, 12:34:33
Job time : 7.04 secs
```

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:26:40 ; Search time 26.18 Seconds
(without alignments)
168.726 Million cell updates/sec

Title: US-09-786-214A-12
Perfect score: 72
Sequence: 1 LPAVVGLSPGEQEY 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	49	68.1	41	11	Q8K408	Q8k408 rattus norv
2	44	61.1	381	16	Q9RRL5	Q9rrl5 deinococcus
3	43.5	60.4	614	16	Q8Y280	Q8y280 ralstonia s
4	43	59.7	821	17	Q9HPR8	Q9hpr8 halobacteri
5	42	58.3	1541	6	Q8HXL3	Q8hxl3 sus scrofa
6	41	56.9	228	17	Q9HKE3	Q9hke3 thermoplasm
7	41	56.9	326	1	Q9UXP0	Q9uxp0 methanolobu
8	41	56.9	579	4	Q8N158	Q8n158 homo sapien
9	41	56.9	877	2	Q7X128	Q7x128 xanthomonas
10	41	56.9	11096	2	Q9L4W3	Q9l4w3 streptomyce
11	40	55.6	154	17	Q8ZYJ3	Q8zyj3 pyrobaculum
12	40	55.6	227	17	Q97WD1	Q97wd1 sulfolobus
13	40	55.6	242	16	Q7WCX1	Q7wcx1 bordetella
14	40	55.6	261	16	Q8NQA2	Q8nqa2 corynebacte
15	40	55.6	298	16	Q7W5D6	Q7w5d6 bordetella
16	40	55.6	298	16	Q7VTI2	Q7vti2 bordetella
17	40	55.6	353	17	Q8ZU64	Q8zu64 pyrobaculum
18	40	55.6	355	10	Q947A7	Q947a7 nitellopsis
19	40	55.6	357	4	Q8NHX1	Q8nhx1 homo sapien
20	40	55.6	358	4	Q9NWD0	Q9nwd0 homo sapien
21	40	55.6	386	4	Q9NWM3	Q9nwm3 homo sapien
22	40	55.6	428	16	Q92AG9	Q92ag9 listeria in
23	40	55.6	428	16	Q8Y661	Q8y661 listeria mo
24	40	55.6	485	10	Q7XNE4	Q7xne4 oryza sativ
25	40	55.6	540	16	Q9RR71	Q9rr71 deinococcus
26	40	55.6	661	10	Q8S8J8	Q8s8j8 arabidopsis
27	40	55.6	665	16	Q8XR57	Q8xr57 ralstonia s
28	40	55.6	683	2	O34302	O34302 rhizobium e
29	40	55.6	753	16	Q89T31	Q89t31 bradyrhizob
30	39	54.2	148	16	Q8F4Y0	Q8f4y0 leptospira
31	39	54.2	197	17	Q8PWI9	Q8pwi9 methanosarc
32	39	54.2	209	17	Q8TTT1	Q8ttt1 methanosarc
33	39	54.2	209	17	Q8TH81	Q8th81 methanosarc
34	39	54.2	277	16	Q9KQK6	Q9kqk6 vibrio chol
35	39	54.2	277	16	Q87RD3	Q87rd3 vibrio para
36	39	54.2	284	16	Q7VFI5	Q7vfi5 helicobacte
37	39	54.2	319	16	Q8ZKQ5	Q8zkq5 salmonella
38	39	54.2	319	16	Q8Z2X4	Q8z2x4 salmonella
39	39	54.2	407	2	Q9LCW0	Q9lcw0 streptomyce
40	39	54.2	589	16	Q7UPH0	Q7uph0 rhodopirell
41	39	54.2	626	16	Q7VWC3	Q7vwc3 bordetella
42	39	54.2	627	16	Q7WH05	Q7wh05 bordetella
43	39	54.2	627	16	Q7W9Q3	Q7w9q3 bordetella
44	39	54.2	1400	11	Q7TME1	Q7tme1 mus musculu
45	38.5	53.5	94	16	Q88F51	Q88f51 pseudomonas

ALIGNMENTS

RESULT 1
Q8K408
ID Q8K408 PRELIMINARY; PRT; 41 AA.
AC Q8K408;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Truncated macrophage colony stimulating factor.
GN CSF1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEW.tl;
RX MEDLINE=22069908; PubMed=12074592;
RA Dobbins D.E., Sood R., Hashiramoto A., Hansen C.T., Wilder R.L.,
RA Remmers E.F.;
RT "Mutation of macrophage colony stimulating factor (Csf1) causes
RT osteopetrosis in the tl rat.";
RL Biochem. Biophys. Res. Commun. 294:1114-1120(2002).
DR EMBL; AF514357; AAM54137.1; -.
SQ SEQUENCE 41 AA; 4178 MW; 1D342C19BD18AA41 CRC64;

Query Match 68.1%; Score 49; DB 11; Length 41;
Best Local Similarity 76.9%; Pred. No. 0.46;
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 LPAVVGLSPGEQE 13
              |||  |||| |||
Db         21 LPAAAGLSPREQE 33
```

RESULT 2
Q9RRL5
ID Q9RRL5 PRELIMINARY; PRT; 381 AA.
AC Q9RRL5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

```
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cytochrome P450.
GN   DR2473.
OS   Deinococcus radiodurans.
OC   Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AE002076; AAF12016.1; -.
DR   PIR; F75270; F75270.
DR   TIGR; DR2473; -.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase; Complete proteome.
SQ   SEQUENCE   381 AA;  41940 MW;  F191EA69F1797B53 CRC64;

  Query Match             61.1%;  Score 44;  DB 16;  Length 381;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSP 9
              |||||||||
Db         52 LPAVVGLSP 60


RESULT 3
Q8Y280
ID   Q8Y280      PRELIMINARY;      PRT;   614 AA.
AC   Q8Y280;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Probable ATP-binding transport ABC transporter protein.
GN   RSC0456 OR RS04444.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Burkholderiaceae; Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646059; CAD13984.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR   GO; GO:0000166; F:nucleotide binding; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR001623; DnaJ_N.
DR   Pfam; PF00005; ABC_tran; 1.
DR   ProDom; PD000006; ABC_transporter; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR   PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
DR   PROSITE; PS00636; DNAJ_1; 1.
KW   Complete proteome.
SQ   SEQUENCE   614 AA;  69240 MW;  E293355B85872142 CRC64;

  Query Match             60.4%;  Score 43.5;  DB 16;  Length 614;
  Best Local Similarity   45.8%;  Pred. No. 76;
  Matches   11;  Conservative    2;  Mismatches    0;  Indels   11;  Gaps    1;

Qy          1 LPAVVG----------LSPGEQE 13
              |||:||          ||||||:
Db        501 LPALVGLLDEVSNWSLRLSPGEQQ 524


RESULT 4
Q9HPR8
ID   Q9HPR8      PRELIMINARY;      PRT;   821 AA.
AC   Q9HPR8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   DNA helicase.
GN   HEL OR VNG1501G.
OS   Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR   EMBL; AE005064; AAG19799.1; -.
DR   PIR; C84304; C84304.
DR   GO; GO:0004386; F:helicase activity; IEA.
DR   GO; GO:0000166; F:nucleotide binding; IEA.
DR   GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001478; PDZ.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00228; PDZ; 1.
KW   Helicase; Complete proteome.
SQ   SEQUENCE   821 AA;  89848 MW;  C454C76B984A5702 CRC64;

  Query Match             59.7%;  Score 43;  DB 17;  Length 821;
  Best Local Similarity   90.0%;  Pred. No. 1.3e+02;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AVVGLSPGEQ 12
              ||||||| ||
Db        326 AVVGLSPAEQ 335


RESULT 5
Q8HXL3
ID   Q8HXL3      PRELIMINARY;      PRT;  1541 AA.
```

```
AC   Q8HXL3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21845885; PubMed=11856879;
RA   Drogemuller C., Kuiper H., Voss-Nemitz R., Brenig B., Distl O.,
RA   Leeb T.;
RT   "Molecular characterization and chromosome assignment of the porcine
RT   gene COX7A1 coding for the muscle specific cytochrome c oxidase
RT   subunit VIIa-M.";
RL   Cytogenet. Cell Genet. 94:190-193(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Leeb T.;
RL   Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ410870; CAD56046.1; -.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00400; WD40; 12.
DR   PROSITE; PS50082; WD_REPEATS_2; 1.
DR   PROSITE; PS50294; WD_REPEATS_REGION; 3.
KW   Hypothetical protein.
SQ   SEQUENCE   1541 AA;  168898 MW;  81B8882854AF4F1E CRC64;

  Query Match             58.3%;  Score 42;  DB 6;  Length 1541;
  Best Local Similarity   58.3%;  Pred. No. 3.7e+02;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQ 12
              :|: : ||||||
Db        786 MPSEISLSPGEQ 797


RESULT 6
Q9HKE3
ID   Q9HKE3      PRELIMINARY;      PRT;   228 AA.
AC   Q9HKE3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Precorrin-2 methyltransferase related protein.
GN   TA0658.
OS   Thermoplasma acidophilum.
OC   Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC   Thermoplasmataceae; Thermoplasma.
OX   NCBI_TaxID=2303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 1728;
RX   MEDLINE=20479972; PubMed=11029001;
RA   Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA   Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT   "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT   acidophilum.";
RL   Nature 407:508-513(2000).
DR   EMBL; AL445065; CAC11796.1; -.
DR   GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. . .; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   GO; GO:0006779; P:porphyrin biosynthesis; IEA.
DR   InterPro; IPR006364; CobI_CbiL.
DR   InterPro; IPR000878; Cor/por_Metransf.
DR   InterPro; IPR003043; Uropor_Metransf.
DR   Pfam; PF00590; TP_methylase; 1.
DR   TIGRFAMs; TIGR01467; cobI_cbiL; 1.
DR   PROSITE; PS00839; SUMT_1; 1.
KW   Transferase; Methyltransferase; Complete proteome.
SQ   SEQUENCE   228 AA;  25084 MW;  11ABD8B68192A67C CRC64;

  Query Match             56.9%;  Score 41;  DB 17;  Length 228;
  Best Local Similarity   63.6%;  Pred. No. 70;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 VVGLSPGEQEY 14
              |||| ||: :|
Db          5 VVGLGPGDPDY 15


RESULT 7
Q9UXP0
ID   Q9UXP0      PRELIMINARY;      PRT;   326 AA.
AC   Q9UXP0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   F420-dependent N5,N10-methylene-tetrahydromethanopterin reductase,
DE   putative.
GN   FFDA.
OS   Methanolobus tindarius.
OC   Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC   Methanosarcinales; Methanosarcinaceae; Methanolobus.
OX   NCBI_TaxID=2221;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 2278;
RA   Westenberg D.J., Braune A., Ruppert C., Mueller V., Herzberg C.,
RA   Gottschalk G., Blaut M.;
RT   "The F420H2-dehydrogenase from Methanolobus tindarius: Cloning of the
RT   ffd operon and expression of the genes in Escherichia coli.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ011519; CAB56639.1; -.
DR   PIR; T45226; T45226.
DR   InterPro; IPR002103; Bac_luciferase.
DR   Pfam; PF00296; bac_luciferase; 1.
SQ   SEQUENCE   326 AA;  34043 MW;  16F3AB9733A45D82 CRC64;

  Query Match             56.9%;  Score 41;  DB 1;  Length 326;
  Best Local Similarity   70.0%;  Pred. No. 1e+02;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AVVGLSPGEQ 12
              |::|| ||||
Db         88 AILGLGPGEQ 97


RESULT 8
Q8N158
ID   Q8N158      PRELIMINARY;      PRT;   579 AA.
AC   Q8N158;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Similar to cerebroglycan (Hypothetical protein FLJ38962).
GN   DKFZP547M109.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA   Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA   Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
```

```
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA   Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Bloecker H., Boecher M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC027972; AAH27972.1; -.
DR   EMBL; AK096281; BAC04745.1; -.
DR   EMBL; AL834418; CAD39080.1; -.
DR   Genew; HGNC:4450; GPC2.
DR   GO; GO:0005578; C:extracellular matrix; IEA.
DR   InterPro; IPR001863; Glypican.
DR   Pfam; PF01153; Glypican; 1.
DR   PROSITE; PS01207; GLYPICAN; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   579 AA;  62829 MW;  1630E3A22BB83DFA CRC64;

  Query Match             56.9%;  Score 41;  DB 4;  Length 579;
  Best Local Similarity   75.0%;  Pred. No. 1.9e+02;
  Matches    9;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQ 12
              || ||| || ||
Db        439 LPPVVGGSPAEQ 450


RESULT 9
Q7X128
ID   Q7X128      PRELIMINARY;      PRT;   877 AA.
AC   Q7X128;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   VirB4.
GN   VIRB4.
OS   Xanthomonas campestris (pv. citri).
OG   Plasmid pXcB.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xanthomonas.
OX   NCBI_TaxID=346;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yuan Q., Brunings A.M., El-Yacoubi B., Shanker S., Gabriel D.W.;
RT   "A self-mobilizing plasmid from a South American citrus canker strain
RT   carries required pathogenicity gene pthB.";
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY228335; AAO72100.1; -.
KW   Plasmid.
SQ   SEQUENCE   877 AA;  100012 MW;  83B94EDB23A0728D CRC64;

  Query Match             56.9%;  Score 41;  DB 2;  Length 877;
  Best Local Similarity   63.6%;  Pred. No. 3e+02;
  Matches    7;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 VVGLSPGEQEY 14
              |:||:||| |:
Db        785 VLGLTPGEYEF 795


RESULT 10
Q9L4W3
ID   Q9L4W3      PRELIMINARY;      PRT; 11096 AA.
AC   Q9L4W3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NysC.
GN   NYSC.
OS   Streptomyces noursei.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11455;
RX   MEDLINE=20334850; PubMed=10873841;
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT   Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT   deduction of the biosynthetic pathway.";
RL   Chem. Biol. 7:395-403(2000).
DR   EMBL; AF263912; AAF71776.1; -.
DR   HSSP; P25715; 1MLA.
DR   GO; GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0004314; F:[acyl-carrier protein] S-malonyltransferase. . .; IEA.
DR   GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR001227; Ac_trans.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl_synth.
DR   InterPro; IPR006162; Ppantne_S.
DR   InterPro; IPR006163; Pp_bind.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00107; ADH_zinc_N; 1.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   11096 AA;  1150415 MW;  776CAEAFCAE551DD CRC64;

  Query Match             56.9%;  Score 41;  DB 2;  Length 11096;
  Best Local Similarity   63.6%;  Pred. No. 4.7e+03;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 PAVVGLSPGEQ 12
              | | ||:||:|
Db       4998 PEVTGLAPGDQ 5008


RESULT 11
Q8ZYJ3
ID   Q8ZYJ3      PRELIMINARY;      PRT;   154 AA.
AC   Q8ZYJ3;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein PAE0746.
GN   PAE0746.
OS   Pyrobaculum aerophilum.
OC   Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC   Thermoproteaceae; Pyrobaculum.
OX   NCBI_TaxID=13773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX   MEDLINE=21664397; PubMed=11792869;
RA   Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA   Miller J.H.;
RT   "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT   aerophilum.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR   EMBL; AE009783; AAL63000.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   154 AA;  17403 MW;  1C3D8BCB40324766 CRC64;
```

```
  Query Match             55.6%;  Score 40;  DB 17;  Length 154;
  Best Local Similarity   66.7%;  Pred. No. 68;
  Matches   10;  Conservative    1;  Mismatches    2;  Indels    2;  Gaps    1;

Qy          1 LPAVVGLS--PGEQE 13
              ||   ||:|  |||||
Db         32 LPPDVGISYTPGEQE 46


RESULT 12
Q97WD1
ID   Q97WD1      PRELIMINARY;      PRT;   227 AA.
AC   Q97WD1;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cobalamin biosynthesis precorrin-6B methylase, putative
DE   (cbiE).
GN   CBIE OR SSO2296.
OS   Sulfolobus solfataricus.
OC   Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC   Sulfolobus.
OX   NCBI_TaxID=2287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35092 / DSM 1617 / P2;
RX   MEDLINE=21332296; PubMed=11427726;
RA   She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA   Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA   De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA   Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA   Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA   Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA   Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT   "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR   EMBL; AE006833; AAK42457.1; -.
DR   PIR; B90400; B90400.
DR   GO; GO:0008168; F:methyltransferase activity; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR000878; Cor/por_Metransf.
DR   Pfam; PF00590; TP_methylase; 1.
KW   Methyltransferase; Complete proteome.
SQ   SEQUENCE   227 AA;  25548 MW;  0267F09491F2DCED CRC64;

  Query Match             55.6%;  Score 40;  DB 17;  Length 227;
  Best Local Similarity   54.5%;  Pred. No. 1e+02;
  Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 VVGLSPGEQEY 14
              :||: ||: ||
Db         10 IVGVGPGDPEY 20


RESULT 13
Q7WCX1
ID   Q7WCX1      PRELIMINARY;      PRT;   242 AA.
AC   Q7WCX1;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Bacteriophage-related DNA polymerase.
GN   BB3808.
OS   Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Bordetella.
OX   NCBI_TaxID=518;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RB50 / ATCC BAA-588;
RX   MEDLINE=22827954; PubMed=12910271;
RA   Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA   Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA   Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA   Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA   Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA   Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA   Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA   Rabbinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA   Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA   Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT   "Comparative analysis of the genome sequences of Bordetella pertussis,
RT   Bordetella parapertussis and Bordetella bronchiseptica.";
RL   Nat. Genet. 35:32-40(2003).
DR   EMBL; BX640448; CAE35782.1; -.
KW   Complete proteome.
SQ   SEQUENCE   242 AA;  25627 MW;  EECAD9B319823284 CRC64;

  Query Match             55.6%;  Score 40;  DB 16;  Length 242;
  Best Local Similarity   80.0%;  Pred. No. 1.1e+02;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 VVGLSPGEQE 13
              ||| :|||||
Db         96 VVGEAPGEQE 105


RESULT 14
Q8NQA2
ID   Q8NQA2      PRELIMINARY;      PRT;   261 AA.
AC   Q8NQA2;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein Cgl1536.
GN   CGL1536.
OS   Corynebacterium glutamicum (Brevibacterium flavum).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX   NCBI_TaxID=1718;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA   Nakagawa S.;
RT   "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP005278; BAB98929.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   261 AA;  27957 MW;  4D8D51D4DCA3A210 CRC64;

  Query Match             55.6%;  Score 40;  DB 16;  Length 261;
  Best Local Similarity   53.8%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQE 13
              |||:  :|||| :
Db        137 LPAIATVSPGEAD 149


RESULT 15
Q7W5D6
ID   Q7W5D6      PRELIMINARY;      PRT;   298 AA.
AC   Q7W5D6;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Bacteriophage-related DNA polymerase.
GN   BPP3357.
OS   Bordetella parapertussis.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Bordetella.
OX   NCBI_TaxID=519;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=12822 / ATCC BAA-587;
```

```
RX   MEDLINE=22827954; PubMed=12910271;
RA   Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA   Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA   Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA   Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA   Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA   Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA   Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA   Rabbinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA   Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA   Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT   "Comparative analysis of the genome sequences of Bordetella pertussis,
RT   Bordetella parapertussis and Bordetella bronchiseptica.";
RL   Nat. Genet. 35:32-40(2003).
DR   EMBL; BX640433; CAE38642.1; -.
KW   Complete proteome.
SQ   SEQUENCE   298 AA;  31202 MW;  A858C8199E2AC8B7 CRC64;

  Query Match             55.6%;  Score 40;  DB 16;  Length 298;
  Best Local Similarity   80.0%;  Pred. No. 1.4e+02;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 VVGLSPGEQE 13
              ||| :|||||
Db        152 VVGEAPGEQE 161


Search completed: May  7, 2004, 12:37:54
Job time : 28.3467 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:19:40 ; Search time 39.06 Seconds
(without alignments)
101.272 Million cell updates/sec

Title: US-09-786-214A-12
Perfect score: 72
Sequence: 1 LPAVVGLSPGEQEY 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	72	100.0	14	3	AAY84266	Aay84266 Peptide d
2	72	100.0	20	3	AAY84265	Aay84265 Truncated
3	72	100.0	25	3	AAY84264	Aay84264 Peptide o
4	68	94.4	13	3	AAY84267	Aay84267 Peptide d
5	65	90.3	13	3	AAY84268	Aay84268 Peptide d
6	65	90.3	15	3	AAY84269	Aay84269 Peptide d
7	45	62.5	234	4	AAB36208	Aab36208 Human imm
8	42	58.3	470	6	AAE34724	Aae34724 Streptomy
9	42	58.3	475	6	AAE34732	Aae34732 Streptomy
10	42	58.3	475	6	AAE34729	Aae34729 Streptomy
11	41	56.9	106	5	ABJ10397	Abj10397 Mutant an
12	41	56.9	106	5	ABJ10395	Abj10395 Mutant an
13	41	56.9	306	4	ABG19552	Abg19552 Novel hum
14	41	56.9	530	7	ADD49105	Add49105 Human NOV
15	41	56.9	549	7	ADD49091	Add49091 Human NOV
16	41	56.9	579	5	ABG70277	Abg70277 Human Gly
17	41	56.9	579	5	ABG97356	Abg97356 Human CGD
18	41	56.9	579	6	ABR39111	Abr39111 Human GPC
19	41	56.9	579	7	ADD49087	Add49087 Human NOV
20	41	56.9	579	7	ADD49107	Add49107 Human NOV
21	41	56.9	579	7	ADD49089	Add49089 Human NOV
22	41	56.9	592	7	ADD49099	Add49099 Human NOV
23	41	56.9	11096	4	AAE10129	Aae10129 Streptomy
24	40	55.6	261	4	AAG91441	Aag91441 C glutami
25	40	55.6	287	6	ABU17217	Abu17217 Protein e
26	40	55.6	303	6	ADA36829	Ada36829 Acinetoba
27	40	55.6	306	4	ABG24698	Abg24698 Novel hum
28	40	55.6	358	4	AAB92530	Aab92530 Human pro
29	40	55.6	359	5	ABB97563	Abb97563 Novel hum
30	40	55.6	428	5	ABB47705	Abb47705 Listeria
31	40	55.6	668	3	AAB57055	Aab57055 Human pro
32	39	54.2	18	2	AAY41875	Aay41875 Rheumatoi
33	39	54.2	18	4	AAU25388	Aau25388 Schizophr
34	39	54.2	18	4	AAU15732	Aau15732 Schizophr
35	39	54.2	18	5	ABG78871	Abg78871 Multiple
36	39	54.2	23	5	ABP62618	Abp62618 Human imm
37	39	54.2	74	6	ABJ18694	Abj18694 Antibody
38	39	54.2	88	3	AAY56655	Aay56655 Partial p
39	39	54.2	93	5	AAU80983	Aau80983 Human ant
40	39	54.2	94	7	ADD69248	Add69248 Human lig
41	39	54.2	95	6	ABO27154	Abo27154 Human ger
42	39	54.2	95	6	ABO27153	Abo27153 Human ger
43	39	54.2	96	6	ABO27155	Abo27155 Human ger
44	39	54.2	96	6	ABO27150	Abo27150 Human ger
45	39	54.2	100	5	AAE23987	Aae23987 Human MOG

ALIGNMENTS

RESULT 1
AAY84266
ID AAY84266 standard; peptide; 14 AA.
XX
AC AAY84266;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Claim 2; Page 39; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying

```
CC   functional variants and mimetics
XX
SQ   Sequence 14 AA;

  Query Match             100.0%;  Score 72;  DB 3;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQEY 14
              ||||||||||||||
Db          1 LPAVVGLSPGEQEY 14


RESULT 2
AAY84265
ID   AAY84265 standard; peptide; 20 AA.
XX
AC   AAY84265;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Truncated macrophage colony stimulating factor tumour antigen.
XX
KW   tumour rejection antigen; macrophage colony stimulating gene;
KW   macrophage-colony stimulating factor; antigen presenting cell;
KW   human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS   Homo sapiens.
XX
PN   WO200013699-A1.
XX
PD   16-MAR-2000.
XX
PF   03-SEP-1999;   99WO-US020344.
XX
PR   04-SEP-1998;   98US-0099077P.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Probst-Kepper M,  Van Den Eynde B,  Boon-Falleur T;
XX
DR   WPI; 2000-256859/22.
DR   N-PSDB; AAZ99675.
XX
PT   Isolated polypeptide used to treat subjects having a disorder
PT   characterized by expression of alternative open reading frame macrophage-
PT   colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS   Claim 3; Page 64; 74pp; English.
XX
CC   The present sequence represents a truncated tumour rejection antigen
CC   precursor, and is encoded by a truncated alternative open reading frame
CC   (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC   the alternative ORF of macrophage-colony stimulating factor, when
CC   presented by an antigen presenting cell having a human leukocyte antigen
CC   (HLA) class I molecule, effectively induce the activation and
CC   proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC   acids derived from the alternate ORF of macrophage-colony stimulating
CC   factor are useful for enriching selectively a population of T lymphocytes
CC   with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC   characterized by expression of the polypeptide, and for identifying
CC   functional variants and mimetics
XX
SQ   Sequence 20 AA;

  Query Match             100.0%;  Score 72;  DB 3;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 9.3e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQEY 14
              ||||||||||||||
Db          4 LPAVVGLSPGEQEY 17


RESULT 3
AAY84264
ID   AAY84264 standard; peptide; 25 AA.
XX
AC   AAY84264;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Peptide of alternate reading frame of macrophage colony stimulating gene.
XX
KW   Renal cell carcinoma; antigen; cytotoxic T lymphocyte;
KW   tumour rejection antigen; macrophage colony stimulating gene;
KW   macrophage-colony stimulating factor; antigen presenting cell;
KW   human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS   Homo sapiens.
XX
PN   WO200013699-A1.
XX
PD   16-MAR-2000.
XX
PF   03-SEP-1999;   99WO-US020344.
XX
PR   04-SEP-1998;   98US-0099077P.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Probst-Kepper M,  Van Den Eynde B,  Boon-Falleur T;
XX
DR   WPI; 2000-256859/22.
DR   N-PSDB; AAZ99672.
XX
PT   Isolated polypeptide used to treat subjects having a disorder
PT   characterized by expression of alternative open reading frame macrophage-
PT   colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS   Claim 1; Page 64; 74pp; English.
XX
CC   The present sequence represents a tumour rejection antigen precursor, and
CC   is encoded by an alternative open reading frame (ORF) of human macrophage
CC   colony stimulating gene. Peptides derived from the alternative ORF of
CC   macrophage-colony stimulating factor, when presented by an antigen
CC   presenting cell having a human leukocyte antigen (HLA) class I molecule,
CC   effectively induce the activation and proliferation of CD8+ cytotoxic T
CC   lymphocytes. Polypeptide and nucleic acids derived from the alternate ORF
CC   of macrophage-colony stimulating factor are useful for enriching
CC   selectively a population of T lymphocytes with CD8+ T lymphocytes. They
CC   are also useful for diagnosing a disorder characterized by expression of
CC   the polypeptide, and for identifying functional variants and mimetics
XX
SQ   Sequence 25 AA;

  Query Match             100.0%;  Score 72;  DB 3;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 0.00012;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQEY 14
              ||||||||||||||
Db          4 LPAVVGLSPGEQEY 17


RESULT 4
AAY84267
ID   AAY84267 standard; peptide; 13 AA.
XX
AC   AAY84267;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW   tumour rejection antigen; macrophage colony stimulating gene;
```

KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Example 2; Page 40; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics
XX
SQ Sequence 13 AA;

Query Match 94.4%; Score 68; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.00026;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQEY 14
||||||||||||
Db 1 PAVVGLSPGEQEY 13

RESULT 5
AAY84268
ID AAY84268 standard; peptide; 13 AA.
XX
AC AAY84268;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Example 2; Page 40; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics
XX
SQ Sequence 13 AA;

Query Match 90.3%; Score 65; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.0008;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPGEQE 13
|||||||||||||
Db 1 LPAVVGLSPGEQE 13

RESULT 6
AAY84269
ID AAY84269 standard; peptide; 15 AA.
XX
AC AAY84269;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX

PS Example 2; Page 40; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics
XX
SQ Sequence 15 AA;

Query Match 90.3%; Score 65; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00094;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPGEQE 13
||||||||||||
Db 3 LPAVVGLSPGEQE 15

RESULT 7
AAB36208
ID AAB36208 standard; protein; 234 AA.
XX
AC AAB36208;
XX
DT 15-FEB-2001 (first entry)
XX
DE Human immune system associated protein HISAP-6.
XX
KW Human; immune system associated protein; HISAP-6; immune disorder;
KW infection; autoimmune disease; cancer.
XX
OS Homo sapiens.
XX
PN US6135941-A.
XX
PD 24-OCT-2000.
XX
PF 27-MAR-1998; 98US-00049672.
XX
PR 27-MAR-1998; 98US-00049672.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Yue H, Lal P, Corley NC, Guegler KJ, Baughn MR;
PI Hillman JL, Au-Young J;
XX
DR WPI; 2001-030926/04.
DR N-PSDB; AAC66524.
XX
PT New human immune system associated proteins (HISAP) and polynucleotides
PT encoding the HISAP, useful for diagnosing, treating or preventing immune
PT or cell proliferative disorders or infections.
XX
PS Claim 1; Col 59-60; 54pp; English.
XX
CC The present invention provides the coding and protein sequences for a
CC number of human immune system associated proteins (HISAPs). These can be
CC used in the diagnosis and treatment of various autoimmune disorders,
CC infections and cell proliferation diseases. The diseases include AIDS,
CC adult respiratory distress syndrome, anaemia, asthma, atherosclerosis,
CC Crohn's disease, irritable bowel syndrome, multiple sclerosis, myasthenia
CC gravis, osteoarthritis, rheumatoid arthritis, scleroderma, systemic lupus
CC erythematosus, arteriosclerosis, cirrhosis and cancer
XX
SQ Sequence 234 AA;

Query Match 62.5%; Score 45; DB 4; Length 234;
Best Local Similarity 72.7%; Pred. No. 36;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQ 12
||:| |||||:
Db 28 PAIVSLSPGER 38

RESULT 8
AAE34724
ID AAE34724 standard; protein; 470 AA.
XX
AC AAE34724;
XX
DT 14-MAY-2003 (first entry)
XX
DE Streptomyces rimosus ema3 protein.
XX
KW P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;
KW emamectin; insecticide; ema3 protein.
XX
OS Streptomyces rimosus.
XX
PN WO200292801-A2.
XX
PD 21-NOV-2002.
XX
PF 15-MAY-2002; 2002WO-EP005363.
XX
PR 16-MAY-2001; 2001US-0291149P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
PI Buckel TG;
XX
DR WPI; 2003-140280/13.
DR N-PSDB; AAD53019.
XX
PT Novel polypeptide that exhibits enzymatic activity of P450 monooxygenase
PT and capable of regioselectively oxidizing alcohol of avermectin useful
PT for making emamectin from avermectin.
XX
PS Claim 17; Page 107-108; 157pp; English.
XX
CC The present invention relates to novel proteins that exhibit an enzymatic
CC activity of P450 monooxygenase and capable of regioselectively oxidising
CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
CC The invention also relates to ferrodoxins and ferrodoxin reductases that
CC are active with the P450 monooxygenases. Sequences of the invention are
CC useful for producing 4-keto-avermectin from avermectin, which is useful
CC for producing emamectin. Emamectin is useful as an insecticide. The
CC present sequence is Streptomyces rimosus ema3 protein
XX
SQ Sequence 470 AA;

Query Match 58.3%; Score 42; DB 6; Length 470;
Best Local Similarity 72.7%; Pred. No. 2.4e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPGE 11
||: ||| |||
Db 44 LPSYVGLHPGE 54

RESULT 9
AAE34732
ID AAE34732 standard; protein; 475 AA.
XX
AC AAE34732;
XX

DT 14-MAY-2003 (first entry)
XX
DE Streptomyces rimosus emal1 protein.
XX
KW P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;
KW emamectin; insecticide; emal1 protein.
XX
OS Streptomyces rimosus.
XX
PN WO200292801-A2.
XX
PD 21-NOV-2002.
XX
PF 15-MAY-2002; 2002WO-EP005363.
XX
PR 16-MAY-2001; 2001US-0291149P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
PI Buckel TG;
XX
DR WPI; 2003-140280/13.
DR N-PSDB; AAD53027.
XX
PT Novel polypeptide that exhibits enzymatic activity of P450 monooxygenase
PT and capable of regioselectively oxidizing alcohol of avermectin useful
PT for making emamectin from avermectin.
XX
PS Claim 17; Page 121-122; 157pp; English.
XX
CC The present invention relates to novel proteins that exhibit an enzymatic
CC activity of P450 monooxygenase and capable of regioselectively oxidising
CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
CC The invention also relates to ferrodoxins and ferrodoxin reductases that
CC are active with the P450 monooxygenases. Sequences of the invention are
CC useful for producing 4-keto-avermectin from avermectin, which is useful
CC for producing emamectin. Emamectin is useful as an insecticide. The
CC present sequence is Streptomyces rimosus emal1 protein
XX
SQ Sequence 475 AA;

Query Match 58.3%; Score 42; DB 6; Length 475;
Best Local Similarity 72.7%; Pred. No. 2.4e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPGE 11
||: ||| |||
Db 49 LPSYVGLHPGE 59

RESULT 10
AAE34729
ID AAE34729 standard; protein; 475 AA.
XX
AC AAE34729;
XX
DT 14-MAY-2003 (first entry)
XX
DE Streptomyces albofaciens ema8 protein.
XX
KW P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;
KW emamectin; insecticide; ema8 protein.
XX
OS Streptomyces albofaciens.
XX
PN WO200292801-A2.
XX
PD 21-NOV-2002.
XX
PF 15-MAY-2002; 2002WO-EP005363.
XX
PR 16-MAY-2001; 2001US-0291149P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
PI Buckel TG;
XX
DR WPI; 2003-140280/13.
DR N-PSDB; AAD53024.
XX
PT Novel polypeptide that exhibits enzymatic activity of P450 monooxygenase
PT and capable of regioselectively oxidizing alcohol of avermectin useful
PT for making emamectin from avermectin.
XX
PS Claim 17; Page 116-117; 157pp; English.
XX
CC The present invention relates to novel proteins that exhibit an enzymatic
CC activity of P450 monooxygenase and capable of regioselectively oxidising
CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
CC The invention also relates to ferrodoxins and ferrodoxin reductases that
CC are active with the P450 monooxygenases. Sequences of the invention are
CC useful for producing 4-keto-avermectin from avermectin, which is useful
CC for producing emamectin. Emamectin is useful as an insecticide. The
CC present sequence is Streptomyces albofaciens ema8 protein
XX
SQ Sequence 475 AA;

Query Match 58.3%; Score 42; DB 6; Length 475;
Best Local Similarity 72.7%; Pred. No. 2.4e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPGE 11
||: ||| |||
Db 49 LPSYVGLHPGE 59

RESULT 11
ABJ10397
ID ABJ10397 standard; protein; 106 AA.
XX
AC ABJ10397;
XX
DT 28-NOV-2002 (first entry)
XX
DE Mutant anti-mesothelin Fv (ST6) variable light chain.
XX
KW Mutant; mutein; recombinant immunotoxin; reduced liver toxicity; cancer.
XX
OS Unidentified.
OS Synthetic.
XX
PN WO200240545-A2.
XX
PD 23-MAY-2002.
XX
PF 16-NOV-2001; 2001WO-US043602.
XX
PR 17-NOV-2000; 2000US-0249805P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Pastan IH, Onda M, Nagata S, Tsutsumi Y, Vincent JJ, Kreitman RJ;
PI Vasmatzis G, Lee B;
XX
DR WPI; 2002-500208/53.
DR N-PSDB; ABT08085.
XX
PT Recombinant immunotoxins with reduced liver toxicity for killing
PT malignant cells bearing antigen comprising mutated framework regions of
PT antibody heavy and/or light chain or antigen-binding portion of
PT immunotoxin.
XX
PS Claim 12; Fig 2; 82pp; English.
XX

```
CC   The invention comprises the amino acid and coding sequences of
CC   recombinant immunotoxin proteins. The immunotoxin proteins of the
CC   invention contain an antibody (or an antigen-binding fragment) with a
CC   substitution of a negatively charged amino acid for an uncharged or
CC   positively charged amino acid. The immunotoxins of the invention have
CC   reduced liver toxicity. The immunotoxins of the invention are useful for
CC   killing a malignant cell (e.g. a cancer cell). The present amino acid
CC   sequence represents a recombinant immunotoxin of the invention
XX
SQ   Sequence 106 AA;

  Query Match             56.9%;  Score 41;  DB 5;  Length 106;
  Best Local Similarity   63.6%;  Pred. No. 67;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 PAVVGLSPGEQ 12
              ||::  |||||
Db          8 PAIMSASPGEQ 18


RESULT 12
ABJ10395
ID   ABJ10395 standard; protein; 106 AA.
XX
AC   ABJ10395;
XX
DT   28-NOV-2002  (first entry)
XX
DE   Mutant anti-Tac Fv (M16) variable light chain.
XX
KW   Mutant; mutein; recombinant immunotoxin; reduced liver toxicity; cancer.
XX
OS   Unidentified.
OS   Synthetic.
XX
PN   WO200240545-A2.
XX
PD   23-MAY-2002.
XX
PF   16-NOV-2001; 2001WO-US043602.
XX
PR   17-NOV-2000; 2000US-0249805P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Pastan IH,  Onda M,  Nagata S,  Tsutsumi Y,  Vincent JJ,  Kreitman RJ;
PI   Vasmatzis G,  Lee B;
XX
DR   WPI; 2002-500208/53.
DR   N-PSDB; ABT08083.
XX
PT   Recombinant immunotoxins with reduced liver toxicity for killing
PT   malignant cells bearing antigen comprising mutated framework regions of
PT   antibody heavy and/or light chain or antigen-binding portion of
PT   immunotoxin.
XX
PS   Claim 8; Fig 2; 82pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of
CC   recombinant immunotoxin proteins. The immunotoxin proteins of the
CC   invention contain an antibody (or an antigen-binding fragment) with a
CC   substitution of a negatively charged amino acid for an uncharged or
CC   positively charged amino acid. The immunotoxins of the invention have
CC   reduced liver toxicity. The immunotoxins of the invention are useful for
CC   killing a malignant cell (e.g. a cancer cell). The present amino acid
CC   sequence represents a recombinant immunotoxin of the invention
XX
SQ   Sequence 106 AA;

  Query Match             56.9%;  Score 41;  DB 5;  Length 106;
  Best Local Similarity   63.6%;  Pred. No. 67;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 PAVVGLSPGEQ 12
              ||::  |||||
Db          8 PAIMSASPGEQ 18


RESULT 13
ABG19552
ID   ABG19552 standard; protein; 306 AA.
XX
AC   ABG19552;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #19543.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS83739.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 49911; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 306 AA;

  Query Match             56.9%;  Score 41;  DB 4;  Length 306;
  Best Local Similarity   57.1%;  Pred. No. 2.2e+02;
  Matches    8;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQEY 14
              ||   ||||| | :
```

```
Db        137 LPQAPGLSPGAQSW 150

RESULT 14
ADD49105
ID   ADD49105 standard; protein; 530 AA.
XX
AC   ADD49105;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Human NOV15k SEQ ID 78.
XX
KW   Antidiabetic; anorectic; cardiant; hypotensive; antiarteriosclerotic;
KW   virucide; antibacterial; fungicide; protozoacide; nootropic;
KW   neuroprotective; antiparkinsonian; anticonvulsant; osteopathic;
KW   antiarthritic; antiinflammatory; dermatological; antiasthmatic;
KW   antilipemic; gene therapy; NOV protein; metabolic disorder; diabetes;
KW   obesity; viral infection; bacterial infection; fungal infection;
KW   helminthic infection; protozoal infection; anorexia; cancer;
KW   cardiovascular disease; hypertension; atherosclerosis;
KW   neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW   epilepsy; immune disorder; osteoarthritis; haematopoietic disorder;
KW   inflammatory skin disorder; asthma; dyslipidemia; human.
XX
OS   Homo sapiens.
XX
PN   WO2003060149-A2.
XX
PD   24-JUL-2003.
XX
PF   06-JAN-2003; 2003WO-US000252.
XX
PR   04-JAN-2002; 2002US-0345222P.
PR   14-JAN-2002; 2002US-0348693P.
PR   16-JAN-2002; 2002US-0349182P.
PR   17-JAN-2002; 2002US-0349733P.
PR   18-JAN-2002; 2002US-0350263P.
PR   24-JAN-2002; 2002US-0351977P.
PR   28-MAY-2002; 2002US-0383758P.
PR   05-JUN-2002; 2002US-0385969P.
PR   11-JUN-2002; 2002US-0387834P.
PR   17-JUL-2002; 2002US-0396407P.
PR   30-SEP-2002; 2002US-0415115P.
PR   03-JAN-2003; 2003US-00336603.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Grosse WM, Alsobrook JP, Anderson DW, Burgess CE, Edinger SR;
PI   Ellerman K, Furtak K, Gangolli EA, Gerlach VL, Gilbert JA;
PI   Gunther E, Gorman L, Guo X, Ji W, Li L, Miller CE, Padigaru M;
PI   Patturajan M, Rastelli L, Macdougall JR, Mishra VS, Smithson G;
PI   Spytek KA, Stone DJ, Shenoy SG, Taupier RJ, Vernet CAM, Zhong M;
PI   Malyankar UM, Millet I, Kekuda R;
XX
DR   WPI; 2003-587288/55.
DR   N-PSDB; ADD49104.
XX
PT   New isolated NOVX polypeptides and polynucleotides, useful for
PT   preventing, diagnosing or treating NOVX-associated disorders, e.g.
PT   osteoarthritis, obesity, atherosclerosis, cancer, Parkinson's disease,
PT   asthma, or infections.
XX
PS   Claim 1; Page 181-182; 311pp; English.
XX
CC   The present invention relates to novel NOV proteins and their coding
CC   sequences (ADD49028-ADD49131). The proteins and coding sequences are
CC   useful in the manufacture of a medicament for treating a syndrome
CC   associated with a human disease, preferably a NOV-associated disorder
CC   such as metabolic disorders, diabetes, obesity, infectious diseases
CC   (viral, bacterial, fungal, helminthic, and protozoal), anorexia, cancer,
CC   cardiovascular diseases (hypertension, atherosclerosis),
CC   neurodegenerative disorders, Alzheimer's disease, Parkinson's disease,
CC   epilepsy, immune disorders (osteoarthritis), hematopoietic disorders,
CC   inflammatory skin disorders, asthma and various dyslipidemias. The coding
CC   sequences and proteins may also be used as targets for the identification
CC   of small molecules that modulate or inhibit e.g. neurogenesis, cell
CC   differentiation, cell proliferation, haematopoiesis, wound healing and
CC   angiogenesis, in gene therapy, in generation of antibodies that bind
CC   immunospecifically to NOV substances for use in therapeutic or diagnostic
CC   methods.
XX
SQ   Sequence 530 AA;

  Query Match             56.9%;  Score 41;  DB 7;  Length 530;
  Best Local Similarity   75.0%;  Pred. No. 4e+02;
  Matches    9;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQ 12
              || ||| || ||
Db        416 LPPVVGGSPAEQ 427


RESULT 15
ADD49091
ID   ADD49091 standard; protein; 549 AA.
XX
AC   ADD49091;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Human NOV15d SEQ ID 64.
XX
KW   Antidiabetic; anorectic; cardiant; hypotensive; antiarteriosclerotic;
KW   virucide; antibacterial; fungicide; protozoacide; nootropic;
KW   neuroprotective; antiparkinsonian; anticonvulsant; osteopathic;
KW   antiarthritic; antiinflammatory; dermatological; antiasthmatic;
KW   antilipemic; gene therapy; NOV protein; metabolic disorder; diabetes;
KW   obesity; viral infection; bacterial infection; fungal infection;
KW   helminthic infection; protozoal infection; anorexia; cancer;
KW   cardiovascular disease; hypertension; atherosclerosis;
KW   neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW   epilepsy; immune disorder; osteoarthritis; haematopoietic disorder;
KW   inflammatory skin disorder; asthma; dyslipidemia; human.
XX
OS   Homo sapiens.
XX
PN   WO2003060149-A2.
XX
PD   24-JUL-2003.
XX
PF   06-JAN-2003; 2003WO-US000252.
XX
PR   04-JAN-2002; 2002US-0345222P.
PR   14-JAN-2002; 2002US-0348693P.
PR   16-JAN-2002; 2002US-0349182P.
PR   17-JAN-2002; 2002US-0349733P.
PR   18-JAN-2002; 2002US-0350263P.
PR   24-JAN-2002; 2002US-0351977P.
PR   28-MAY-2002; 2002US-0383758P.
PR   05-JUN-2002; 2002US-0385969P.
PR   11-JUN-2002; 2002US-0387834P.
PR   17-JUL-2002; 2002US-0396407P.
PR   30-SEP-2002; 2002US-0415115P.
PR   03-JAN-2003; 2003US-00336603.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Grosse WM, Alsobrook JP, Anderson DW, Burgess CE, Edinger SR;
PI   Ellerman K, Furtak K, Gangolli EA, Gerlach VL, Gilbert JA;
PI   Gunther E, Gorman L, Guo X, Ji W, Li L, Miller CE, Padigaru M;
PI   Patturajan M, Rastelli L, Macdougall JR, Mishra VS, Smithson G;
PI   Spytek KA, Stone DJ, Shenoy SG, Taupier RJ, Vernet CAM, Zhong M;
PI   Malyankar UM, Millet I, Kekuda R;
XX
DR   WPI; 2003-587288/55.
```

DR N-PSDB; ADD49090.
XX
PT New isolated NOVX polypeptides and polynucleotides, useful for
PT preventing, diagnosing or treating NOVX-associated disorders, e.g.
PT osteoarthritis, obesity, atherosclerosis, cancer, Parkinson's disease,
PT asthma, or infections.
XX
PS Claim 1; Page 177; 311pp; English.
XX
CC The present invention relates to novel NOV proteins and their coding
CC sequences (ADD49028-ADD49131). The proteins and coding sequences are
CC useful in the manufacture of a medicament for treating a syndrome
CC associated with a human disease, preferably a NOV-associated disorder
CC such as metabolic disorders, diabetes, obesity, infectious diseases
CC (viral, bacterial, fungal, helminthic, and protozoal), anorexia, cancer,
CC cardiovascular diseases (hypertension, atherosclerosis),
CC neurodegenerative disorders, Alzheimer's disease, Parkinson's disease,
CC epilepsy, immune disorders (osteoarthritis), hematopoietic disorders,
CC inflammatory skin disorders, asthma and various dyslipidemias. The coding
CC sequences and proteins may also be used as targets for the identification
CC of small molecules that modulate or inhibit e.g. neurogenesis, cell
CC differentiation, cell proliferation, haematopoiesis, wound healing and
CC angiogenesis, in gene therapy, in generation of antibodies that bind
CC immunospecifically to NOV substances for use in therapeutic or diagnostic
CC methods.
XX
SQ Sequence 549 AA;

Query Match 56.9%; Score 41; DB 7; Length 549;
Best Local Similarity 75.0%; Pred. No. 4.1e+02;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 LPAVVGLSPGEQ 12
              || ||| || ||
Db        419 LPPVVGGSPAEQ 430
```

Search completed: May 7, 2004, 12:33:45
Job time : 41.06 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:26:40 ; Search time 24.31 Seconds
(without alignments)
168.726 Million cell updates/sec

Title: US-09-786-214A-13
Perfect score: 68
Sequence: 1 PAVVGLSPGEQEY 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	45	66.2	41	11	Q8K408	Q8k408 rattus norv
2	43	63.2	821	17	Q9HPR8	Q9hpr8 halobacteri
3	41	60.3	228	17	Q9HKE3	Q9hke3 thermoplasm
4	41	60.3	326	1	Q9UXP0	Q9uxp0 methanolobu
5	41	60.3	877	2	Q7X128	Q7x128 xanthomonas
6	41	60.3	11096	2	Q9L4W3	Q9l4w3 streptomyce
7	40	58.8	227	17	Q97WD1	Q97wd1 sulfolobus
8	40	58.8	242	16	Q7WCX1	Q7wcx1 bordetella
9	40	58.8	298	16	Q7W5D6	Q7w5d6 bordetella
10	40	58.8	298	16	Q7VTI2	Q7vti2 bordetella
11	40	58.8	355	10	Q947A7	Q947a7 nitellopsis
12	40	58.8	357	4	Q8NHX1	Q8nhx1 homo sapien
13	40	58.8	358	4	Q9NWD0	Q9nwd0 homo sapien
14	40	58.8	381	16	Q9RRL5	Q9rrl5 deinococcus
15	40	58.8	386	4	Q9NWM3	Q9nwm3 homo sapien
16	40	58.8	540	16	Q9RR71	Q9rr71 deinococcus
17	40	58.8	661	10	Q8S8J8	Q8s8j8 arabidopsis
18	40	58.8	665	16	Q8XR57	Q8xr57 ralstonia s
19	40	58.8	683	2	O34302	O34302 rhizobium e
20	40	58.8	1541	6	Q8HXL3	Q8hxl3 sus scrofa
21	39.5	58.1	614	16	Q8Y280	Q8y280 ralstonia s
22	39	57.4	148	16	Q8F4Y0	Q8f4y0 leptospira
23	39	57.4	197	17	Q8PWI9	Q8pwi9 methanosarc
24	39	57.4	209	17	Q8TTT1	Q8ttt1 methanosarc
25	39	57.4	209	17	Q8TH81	Q8th81 methanosarc
26	39	57.4	407	2	Q9LCW0	Q9lcw0 streptomyce
27	39	57.4	428	16	Q92AG9	Q92ag9 listeria in
28	39	57.4	428	16	Q8Y661	Q8y661 listeria mo
29	39	57.4	589	16	Q7UPH0	Q7uph0 rhodopirell
30	39	57.4	626	16	Q7VWC3	Q7vwc3 bordetella
31	39	57.4	627	16	Q7WH05	Q7wh05 bordetella
32	39	57.4	627	16	Q7W9Q3	Q7w9q3 bordetella
33	38.5	56.6	94	16	Q88F51	Q88f51 pseudomonas
34	38	55.9	136	11	Q61061	Q61061 mus musculu
35	38	55.9	152	17	Q9V2S2	Q9v2s2 pyrococcus
36	38	55.9	152	17	Q8TZK1	Q8tzk1 pyrococcus
37	38	55.9	155	17	O57778	O57778 pyrococcus
38	38	55.9	200	16	Q97LH7	Q97lh7 clostridium
39	38	55.9	203	16	Q8XL17	Q8xl17 clostridium
40	38	55.9	222	17	Q97WC8	Q97wc8 sulfolobus
41	38	55.9	313	5	Q7YTA2	Q7yta2 glomeris ma
42	38	55.9	325	16	Q9CBS2	Q9cbs2 mycobacteri
43	38	55.9	352	16	Q8G7R2	Q8g7r2 bifidobacte
44	38	55.9	353	17	Q8ZU64	Q8zu64 pyrobaculum
45	38	55.9	404	16	Q897K0	Q897k0 clostridium

ALIGNMENTS

RESULT 1
Q8K408
ID Q8K408 PRELIMINARY; PRT; 41 AA.
AC Q8K408;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Truncated macrophage colony stimulating factor.
GN CSF1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEW.tl;
RX MEDLINE=22069908; PubMed=12074592;
RA Dobbins D.E., Sood R., Hashiramoto A., Hansen C.T., Wilder R.L.,
RA Remmers E.F.;
RT "Mutation of macrophage colony stimulating factor (Csf1) causes
RT osteopetrosis in the tl rat.";
RL Biochem. Biophys. Res. Commun. 294:1114-1120(2002).
DR EMBL; AF514357; AAM54137.1; -.
SQ SEQUENCE 41 AA; 4178 MW; 1D342C19BD18AA41 CRC64;

Query Match 66.2%; Score 45; DB 11; Length 41;
Best Local Similarity 75.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 PAVVGLSPGEQE 12
              ||  |||| |||
Db         22 PAAAGLSPREQE 33
```

RESULT 2
Q9HPR8
ID Q9HPR8 PRELIMINARY; PRT; 821 AA.
AC Q9HPR8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)

```
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   DNA helicase.
GN   HEL OR VNG1501G.
OS   Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR   EMBL; AE005064; AAG19799.1; -.
DR   PIR; C84304; C84304.
DR   GO; GO:0004386; F:helicase activity; IEA.
DR   GO; GO:0000166; F:nucleotide binding; IEA.
DR   GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001478; PDZ.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00228; PDZ; 1.
KW   Helicase; Complete proteome.
SQ   SEQUENCE   821 AA;  89848 MW;  C454C76B984A5702 CRC64;

  Query Match             63.2%;  Score 43;  DB 17;  Length 821;
  Best Local Similarity   90.0%;  Pred. No. 91;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 AVVGLSPGEQ 11
              |||||||  ||
Db        326 AVVGLSPAEQ 335


RESULT 3
Q9HKE3
ID   Q9HKE3      PRELIMINARY;      PRT;   228 AA.
AC   Q9HKE3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Precorrin-2 methyltransferase related protein.
GN   TA0658.
OS   Thermoplasma acidophilum.
OC   Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC   Thermoplasmataceae; Thermoplasma.
OX   NCBI_TaxID=2303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 1728;
RX   MEDLINE=20479972; PubMed=11029001;
RA   Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA   Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT   "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT   acidophilum.";
RL   Nature 407:508-513(2000).
DR   EMBL; AL445065; CAC11796.1; -.
DR   GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. . .; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   GO; GO:0006779; P:porphyrin biosynthesis; IEA.
DR   InterPro; IPR006364; CobI_CbiL.
DR   InterPro; IPR000878; Cor/por_Metransf.
DR   InterPro; IPR003043; Uropor_Metransf.
DR   Pfam; PF00590; TP_methylase; 1.
DR   TIGRFAMs; TIGR01467; cobI_cbiL; 1.
DR   PROSITE; PS00839; SUMT_1; 1.
KW   Transferase; Methyltransferase; Complete proteome.
SQ   SEQUENCE   228 AA;  25084 MW;  11ABD8B68192A67C CRC64;

  Query Match             60.3%;  Score 41;  DB 17;  Length 228;
  Best Local Similarity   63.6%;  Pred. No. 51;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 VVGLSPGEQEY 13
              ||||  ||: :|
Db          5 VVGLGPGDPDY 15


RESULT 4
Q9UXP0
ID   Q9UXP0      PRELIMINARY;      PRT;   326 AA.
AC   Q9UXP0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   F420-dependent N5,N10-methylene-tetrahydromethanopterin reductase,
DE   putative.
GN   FFDA.
OS   Methanolobus tindarius.
OC   Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC   Methanosarcinales; Methanosarcinaceae; Methanolobus.
OX   NCBI_TaxID=2221;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 2278;
RA   Westenberg D.J., Braune A., Ruppert C., Mueller V., Herzberg C.,
RA   Gottschalk G., Blaut M.;
RT   "The F420H2-dehydrogenase from Methanolobus tindarius: Cloning of the
RT   ffd operon and expression of the genes in Escherichia coli.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ011519; CAB56639.1; -.
DR   PIR; T45226; T45226.
DR   InterPro; IPR002103; Bac_luciferase.
DR   Pfam; PF00296; bac_luciferase; 1.
SQ   SEQUENCE   326 AA;  34043 MW;  16F3AB9733A45D82 CRC64;

  Query Match             60.3%;  Score 41;  DB 1;  Length 326;
  Best Local Similarity   70.0%;  Pred. No. 75;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 AVVGLSPGEQ 11
              |::||  ||||
Db         88 AILGLGPGEQ 97


RESULT 5
Q7X128
ID   Q7X128      PRELIMINARY;      PRT;   877 AA.
AC   Q7X128;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   VirB4.
GN   VIRB4.
OS   Xanthomonas campestris (pv. citri).
OG   Plasmid pXcB.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xanthomonas.
OX   NCBI_TaxID=346;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yuan Q., Brunings A.M., El-Yacoubi B., Shanker S., Gabriel D.W.;
RT   "A self-mobilizing plasmid from a South American citrus canker strain
RT   carries required pathogenicity gene pthB.";
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY228335; AAO72100.1; -.
KW   Plasmid.
```

```
SQ   SEQUENCE   877 AA;  100012 MW;  83B94EDB23A0728D CRC64;

  Query Match             60.3%;  Score 41;  DB 2;  Length 877;
  Best Local Similarity   63.6%;  Pred. No. 2.2e+02;
  Matches    7;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 VVGLSPGEQEY 13
              |:||:||| |:
Db        785 VLGLTPGEYEF 795


RESULT 6
Q9L4W3
ID   Q9L4W3      PRELIMINARY;      PRT; 11096 AA.
AC   Q9L4W3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NysC.
GN   NYSC.
OS   Streptomyces noursei.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11455;
RX   MEDLINE=20334850; PubMed=10873841;
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT   Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT   deduction of the biosynthetic pathway.";
RL   Chem. Biol. 7:395-403(2000).
DR   EMBL; AF263912; AAF71776.1; -.
DR   HSSP; P25715; 1MLA.
DR   GO; GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0004314; F:[acyl-carrier protein] S-malonyltransferase. . .; IEA.
DR   GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR001227; Ac_trans.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl_synth.
DR   InterPro; IPR006162; Ppantne_S.
DR   InterPro; IPR006163; Pp_bind.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00107; ADH_zinc_N; 1.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   11096 AA;  1150415 MW;  776CAEAFCAE551DD CRC64;

  Query Match             60.3%;  Score 41;  DB 2;  Length 11096;
  Best Local Similarity   63.6%;  Pred. No. 3.4e+03;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQ 11
              | | ||:||:|
Db       4998 PEVTGLAPGDQ 5008


RESULT 7
Q97WD1
ID   Q97WD1      PRELIMINARY;      PRT;  227 AA.
AC   Q97WD1;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cobalamin biosynthesis precorrin-6B methylase, putative
DE   (cbiE).
GN   CBIE OR SSO2296.
OS   Sulfolobus solfataricus.
OC   Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC   Sulfolobus.
OX   NCBI_TaxID=2287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35092 / DSM 1617 / P2;
RX   MEDLINE=21332296; PubMed=11427726;
RA   She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA   Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA   De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA   Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA   Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA   Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA   Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT   "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR   EMBL; AE006833; AAK42457.1; -.
DR   PIR; B90400; B90400.
DR   GO; GO:0008168; F:methyltransferase activity; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR000878; Cor/por_Metransf.
DR   Pfam; PF00590; TP_methylase; 1.
KW   Methyltransferase; Complete proteome.
SQ   SEQUENCE   227 AA;  25548 MW;  0267F09491F2DCED CRC64;

  Query Match             58.8%;  Score 40;  DB 17;  Length 227;
  Best Local Similarity   54.5%;  Pred. No. 76;
  Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 VVGLSPGEQEY 13
              :||: ||: ||
Db         10 IVGVGPGDPEY 20


RESULT 8
Q7WCX1
ID   Q7WCX1      PRELIMINARY;      PRT;  242 AA.
AC   Q7WCX1;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Bacteriophage-related DNA polymerase.
GN   BB3808.
OS   Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Bordetella.
OX   NCBI_TaxID=518;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RB50 / ATCC BAA-588;
RX   MEDLINE=22827954; PubMed=12910271;
RA   Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA   Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA   Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA   Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA   Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA   Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA   Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA   Rabbinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA   Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA   Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT   "Comparative analysis of the genome sequences of Bordetella pertussis,
RT   Bordetella parapertussis and Bordetella bronchiseptica.";
RL   Nat. Genet. 35:32-40(2003).
DR   EMBL; BX640448; CAE35782.1; -.
KW   Complete proteome.
```

```
SQ   SEQUENCE   242 AA;  25627 MW;  EECAD9B319823284 CRC64;

  Query Match             58.8%;  Score 40;  DB 16;  Length 242;
  Best Local Similarity   80.0%;  Pred. No. 82;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 VVGLSPGEQE 12
              ||| :|||||
Db         96 VVGEAPGEQE 105


RESULT 9
Q7W5D6
ID   Q7W5D6      PRELIMINARY;      PRT;   298 AA.
AC   Q7W5D6;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Bacteriophage-related DNA polymerase.
GN   BPP3357.
OS   Bordetella parapertussis.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Bordetella.
OX   NCBI_TaxID=519;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=12822 / ATCC BAA-587;
RX   MEDLINE=22827954; PubMed=12910271;
RA   Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA   Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA   Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA   Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA   Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA   Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA   Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA   Rabbinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA   Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA   Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT   "Comparative analysis of the genome sequences of Bordetella pertussis,
RT   Bordetella parapertussis and Bordetella bronchiseptica.";
RL   Nat. Genet. 35:32-40(2003).
DR   EMBL; BX640433; CAE38642.1; -.
KW   Complete proteome.
SQ   SEQUENCE   298 AA;  31202 MW;  A858C8199E2AC8B7 CRC64;

  Query Match             58.8%;  Score 40;  DB 16;  Length 298;
  Best Local Similarity   80.0%;  Pred. No. 1e+02;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 VVGLSPGEQE 12
              ||| :|||||
Db        152 VVGEAPGEQE 161


RESULT 10
Q7VTI2
ID   Q7VTI2      PRELIMINARY;      PRT;   298 AA.
AC   Q7VTI2;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Bacteriophage-related DNA polymerase.
GN   BP3556.
OS   Bordetella pertussis.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Bordetella.
OX   NCBI_TaxID=520;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RX   MEDLINE=22827954; PubMed=12910271;
RA   Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA   Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA   Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA   Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA   Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA   Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA   Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA   Rabbinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA   Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA   Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT   "Comparative analysis of the genome sequences of Bordetella pertussis,
RT   Bordetella parapertussis and Bordetella bronchiseptica.";
RL   Nat. Genet. 35:32-40(2003).
DR   EMBL; BX640421; CAE43815.1; -.
KW   Complete proteome.
SQ   SEQUENCE   298 AA;  31231 MW;  9E4BC84B419506F3 CRC64;

  Query Match             58.8%;  Score 40;  DB 16;  Length 298;
  Best Local Similarity   80.0%;  Pred. No. 1e+02;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 VVGLSPGEQE 12
              ||| :|||||
Db        152 VVGEAPGEQE 161


RESULT 11
Q947A7
ID   Q947A7      PRELIMINARY;      PRT;   355 AA.
AC   Q947A7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Enolase (EC 4.2.1.11) (2-phosphoglycerate dehydratase) (2-phospho-D-
DE   glycerate hydro-lyase) (Fragment).
GN   ENO.
OS   Nitellopsis obtusa.
OC   Eukaryota; Viridiplantae; Streptophyta; Charales; Characeae;
OC   Nitellopsis.
OX   NCBI_TaxID=40811;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=X-854;
RX   MEDLINE=21437986; PubMed=11526220;
RA   Keeling P.J., Palmer J.D.;
RT   "Lateral transfer at the gene and subgenic levels in the evolution of
RT   eukaryotic enolase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:10745-10750(2001).
CC   -!- CATALYTIC ACTIVITY: 2-PHOSPHO-D-GLYCERATE = PHOSPHOENOLPYRUVATE +
CC       H(2)O.
CC   -!- COFACTOR: MAGNESIUM IS REQUIRED FOR CATALYSIS AND FOR STABILIZING
CC       THE DIMER (BY SIMILARITY).
CC   -!- PATHWAY: GLYCOLYSIS.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ENOLASE FAMILY.
DR   EMBL; AF348916; AAL05455.1; -.
DR   GO; GO:0000015; C:phosphopyruvate hydratase complex; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0000287; F:magnesium ion binding; IEA.
DR   GO; GO:0004634; F:phosphopyruvate hydratase activity; IEA.
DR   GO; GO:0006096; P:glycolysis; IEA.
DR   InterPro; IPR000941; Enolase.
DR   Pfam; PF00113; enolase; 1.
DR   Pfam; PF03952; enolase_N; 1.
DR   PRINTS; PR00148; ENOLASE.
DR   ProDom; PD000902; Enolase; 1.
DR   PROSITE; PS00164; ENOLASE; 1.
KW   Glycolysis; Lyase; Magnesium.
FT   NON_TER       1      1
FT   NON_TER     355    355
SQ   SEQUENCE   355 AA;  38295 MW;  400DF160087DE450 CRC64;

  Query Match             58.8%;  Score 40;  DB 10;  Length 355;
  Best Local Similarity   54.5%;  Pred. No. 1.2e+02;
```

```
  Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQ 11
              |||:|: | :|
Db         28 PAVIGMDPADQ 38


RESULT 12
Q8NHX1
ID   Q8NHX1      PRELIMINARY;      PRT;   357 AA.
AC   Q8NHX1;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Extracellular signal-related kinase 1b.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Aebersold D.M., Yung Y., Seger R.;
RT   "Properties of human ERK1b.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY033607; AAK52329.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004707; F:MAP kinase activity; IEA.
DR   GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR   GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   InterPro; IPR008349; Erk_1_2_MAPK.
DR   InterPro; IPR008350; Erk_3_4_MAPK.
DR   InterPro; IPR003527; MAP_kin.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR01770; ERK1ERK2MAPK.
DR   PRINTS; PR01771; ERK3ERK4MAPK.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS01351; MAPK; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   ATP-binding; Kinase; Transferase.
SQ   SEQUENCE   357 AA;  40062 MW;  58C92773983ADA79 CRC64;

  Query Match             58.8%;  Score 40;  DB 4;  Length 357;
  Best Local Similarity   72.7%;  Pred. No. 1.2e+02;
  Matches    8;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQ 11
              || |||   |||
Db        344 PAAVGLGAGEQ 354


RESULT 13
Q9NWD0
ID   Q9NWD0      PRELIMINARY;      PRT;   358 AA.
AC   Q9NWD0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA   Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA   Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA   Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA   Ninomiya K., Iwayanagi T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK000977; BAA91452.1; -.
DR   InterPro; IPR003892; CUE.
DR   Pfam; PF02845; CUE; 1.
DR   SMART; SM00546; CUE; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   358 AA;  38760 MW;  BCE1AA7E95C73BF0 CRC64;

  Query Match             58.8%;  Score 40;  DB 4;  Length 358;
  Best Local Similarity   58.3%;  Pred. No. 1.3e+02;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQE 12
              ||: |  ||:||
Db        213 PAMAGPGPGDQE 224


RESULT 14
Q9RRL5
ID   Q9RRL5      PRELIMINARY;      PRT;   381 AA.
AC   Q9RRL5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cytochrome P450.
GN   DR2473.
OS   Deinococcus radiodurans.
OC   Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AE002076; AAF12016.1; -.
DR   PIR; F75270; F75270.
DR   TIGR; DR2473; -.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase; Complete proteome.
SQ   SEQUENCE   381 AA;  41940 MW;  F191EA69F1797B53 CRC64;

  Query Match             58.8%;  Score 40;  DB 16;  Length 381;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSP 8
              ||||||||
Db         53 PAVVGLSP 60
```

```
RESULT 15
Q9NWM3
ID   Q9NWM3      PRELIMINARY;      PRT;   386 AA.
AC   Q9NWM3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein FLJ20739.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA   Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,
RA   Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK000746; BAA91357.1; -.
DR   Genew; HGNC:19843; C14orf34.
DR   InterPro; IPR003892; CUE.
DR   Pfam; PF02845; CUE; 1.
DR   SMART; SM00546; CUE; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   386 AA;  42258 MW;  F5530FA47C267895 CRC64;

  Query Match             58.8%;  Score 40;  DB 4;  Length 386;
  Best Local Similarity   58.3%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQE 12
              ||: |  ||:||
Db        213 PAMAGPGPGDQE 224


Search completed: May  7, 2004, 12:37:55
Job time : 25.4767 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:19:40 ; Search time 36.27 Seconds
(without alignments)
101.272 Million cell updates/sec

Title: US-09-786-214A-13
Perfect score: 68
Sequence: 1 PAVVGLSPGEQEY 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	68	100.0	13	3	AAY84267	Aay84267 Peptide d
2	68	100.0	14	3	AAY84266	Aay84266 Peptide d
3	68	100.0	20	3	AAY84265	Aay84265 Truncated
4	68	100.0	25	3	AAY84264	Aay84264 Peptide o
5	61	89.7	13	3	AAY84268	Aay84268 Peptide d
6	61	89.7	15	3	AAY84269	Aay84269 Peptide d
7	45	66.2	234	4	AAB36208	Aab36208 Human imm
8	41	60.3	106	5	ABJ10397	Abj10397 Mutant an
9	41	60.3	106	5	ABJ10395	Abj10395 Mutant an
10	41	60.3	11096	4	AAE10129	Aae10129 Streptomy
11	40	58.8	358	4	AAB92530	Aab92530 Human pro
12	40	58.8	359	5	ABB97563	Abb97563 Novel hum
13	40	58.8	668	3	AAB57055	Aab57055 Human pro
14	39	57.4	18	2	AAY41875	Aay41875 Rheumatoi
15	39	57.4	18	4	AAU25388	Aau25388 Schizophr
16	39	57.4	18	4	AAU15732	Aau15732 Schizophr
17	39	57.4	18	5	ABG78871	Abg78871 Multiple
18	39	57.4	23	5	ABP62618	Abp62618 Human imm
19	39	57.4	74	6	ABJ18694	Abj18694 Antibody
20	39	57.4	88	3	AAY56655	Aay56655 Partial p
21	39	57.4	93	5	AAU80983	Aau80983 Human ant
22	39	57.4	94	7	ADD69248	Add69248 Human lig
23	39	57.4	95	6	ABO27154	Abo27154 Human ger
24	39	57.4	95	6	ABO27153	Abo27153 Human ger
25	39	57.4	96	6	ABO27155	Abo27155 Human ger
26	39	57.4	96	6	ABO27150	Abo27150 Human ger
27	39	57.4	100	5	AAE23987	Aae23987 Human MOG
28	39	57.4	104	2	AAW31723	Aaw31723 Alpha lig
29	39	57.4	104	2	AAW26795	Aaw26795 Anti-gp54
30	39	57.4	105	7	ABR61567	Abr61567 HIV-1 neu
31	39	57.4	106	2	AAR41234	Aar41234 Monoclona
32	39	57.4	106	2	AAW71241	Aaw71241 Light cha
33	39	57.4	106	6	AAE34874	Aae34874 BIWA4 ant
34	39	57.4	106	6	AAE34875	Aae34875 BIWA8 ant
35	39	57.4	106	6	AAE33419	Aae33419 Murine KS
36	39	57.4	106	6	AAE33421	Aae33421 Murine KS
37	39	57.4	106	6	AAO30909	Aao30909 hu-KS ant
38	39	57.4	107	2	AAR32129	Aar32129 Anti-IL2R
39	39	57.4	107	2	AAR37610	Aar37610 B-B10 MAb
40	39	57.4	107	2	AAR50190	Aar50190 Light cha
41	39	57.4	107	2	AAW08949	Aaw08949 Kappa lig
42	39	57.4	107	2	AAW08948	Aaw08948 Kappa lig
43	39	57.4	107	2	AAW24514	Aaw24514 Anti-Fact
44	39	57.4	107	2	AAW24515	Aaw24515 Anti-Fact
45	39	57.4	107	2	AAW24513	Aaw24513 Anti-Fact

ALIGNMENTS

RESULT 1
AAY84267
ID AAY84267 standard; peptide; 13 AA.
XX
AC AAY84267;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Example 2; Page 40; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying

```
CC   functional variants and mimetics
XX
SQ   Sequence 13 AA;

  Query Match             100.0%;  Score 68;  DB 3;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 0.00015;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQEY 13
              |||||||||||||
Db          1 PAVVGLSPGEQEY 13


RESULT 2
AAY84266
ID   AAY84266 standard; peptide; 14 AA.
XX
AC   AAY84266;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW   tumour rejection antigen; macrophage colony stimulating gene;
KW   macrophage-colony stimulating factor; antigen presenting cell;
KW   human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO200013699-A1.
XX
PD   16-MAR-2000.
XX
PF   03-SEP-1999;   99WO-US020344.
XX
PR   04-SEP-1998;   98US-0099077P.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Probst-Kepper M,  Van Den Eynde B,  Boon-Falleur T;
XX
DR   WPI; 2000-256859/22.
XX
PT   Isolated polypeptide used to treat subjects having a disorder
PT   characterized by expression of alternative open reading frame macrophage-
PT   colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS   Claim 2; Page 39; 74pp; English.
XX
CC   The present sequence represents a peptide which is derived from a tumour
CC   rejection antigen precursor encoded by an alternative open reading frame
CC   (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC   the alternative ORF of macrophage-colony stimulating factor, when
CC   presented by an antigen presenting cell having a human leukocyte antigen
CC   (HLA) class I molecule, effectively induce the activation and
CC   proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC   acids derived from the alternate ORF of macrophage-colony stimulating
CC   factor are useful for enriching selectively a population of T lymphocytes
CC   with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC   characterized by expression of the polypeptide, and for identifying
CC   functional variants and mimetics
XX
SQ   Sequence 14 AA;

  Query Match             100.0%;  Score 68;  DB 3;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.00016;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQEY 13
              |||||||||||||
Db          2 PAVVGLSPGEQEY 14


RESULT 3
AAY84265
ID   AAY84265 standard; peptide; 20 AA.
XX
AC   AAY84265;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Truncated macrophage colony stimulating factor tumour antigen.
XX
KW   tumour rejection antigen; macrophage colony stimulating gene;
KW   macrophage-colony stimulating factor; antigen presenting cell;
KW   human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS   Homo sapiens.
XX
PN   WO200013699-A1.
XX
PD   16-MAR-2000.
XX
PF   03-SEP-1999;   99WO-US020344.
XX
PR   04-SEP-1998;   98US-0099077P.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Probst-Kepper M,  Van Den Eynde B,  Boon-Falleur T;
XX
DR   WPI; 2000-256859/22.
DR   N-PSDB; AAZ99675.
XX
PT   Isolated polypeptide used to treat subjects having a disorder
PT   characterized by expression of alternative open reading frame macrophage-
PT   colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS   Claim 3; Page 64; 74pp; English.
XX
CC   The present sequence represents a truncated tumour rejection antigen
CC   precursor, and is encoded by a truncated alternative open reading frame
CC   (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC   the alternative ORF of macrophage-colony stimulating factor, when
CC   presented by an antigen presenting cell having a human leukocyte antigen
CC   (HLA) class I molecule, effectively induce the activation and
CC   proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC   acids derived from the alternate ORF of macrophage-colony stimulating
CC   factor are useful for enriching selectively a population of T lymphocytes
CC   with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC   characterized by expression of the polypeptide, and for identifying
CC   functional variants and mimetics
XX
SQ   Sequence 20 AA;

  Query Match             100.0%;  Score 68;  DB 3;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 0.00024;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQEY 13
              |||||||||||||
Db          5 PAVVGLSPGEQEY 17


RESULT 4
AAY84264
ID   AAY84264 standard; peptide; 25 AA.
XX
AC   AAY84264;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Peptide of alternate reading frame of macrophage colony stimulating gene.
XX
KW   Renal cell carcinoma; antigen; cytotoxic T lymphocyte;
```

KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
DR N-PSDB; AAZ99672.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Claim 1; Page 64; 74pp; English.
XX
CC The present sequence represents a tumour rejection antigen precursor, and
CC is encoded by an alternative open reading frame (ORF) of human macrophage
CC colony stimulating gene. Peptides derived from the alternative ORF of
CC macrophage-colony stimulating factor, when presented by an antigen
CC presenting cell having a human leukocyte antigen (HLA) class I molecule,
CC effectively induce the activation and proliferation of CD8+ cytotoxic T
CC lymphocytes. Polypeptide and nucleic acids derived from the alternate ORF
CC of macrophage-colony stimulating factor are useful for enriching
CC selectively a population of T lymphocytes with CD8+ T lymphocytes. They
CC are also useful for diagnosing a disorder characterized by expression of
CC the polypeptide, and for identifying functional variants and mimetics
XX
SQ Sequence 25 AA;

Query Match 100.0%; Score 68; DB 3; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.00031;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PAVVGLSPGEQEY 13
|||||||||||||
Db 5 PAVVGLSPGEQEY 17

RESULT 5
AAY84268
ID AAY84268 standard; peptide; 13 AA.
XX
AC AAY84268;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Example 2; Page 40; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics
XX
SQ Sequence 13 AA;

Query Match 89.7%; Score 61; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.0022;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PAVVGLSPGEQE 12
||||||||||||
Db 2 PAVVGLSPGEQE 13

RESULT 6
AAY84269
ID AAY84269 standard; peptide; 15 AA.
XX
AC AAY84269;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX

```
PS   Example 2; Page 40; 74pp; English.
XX
CC   The present sequence represents a peptide which is derived from a tumour
CC   rejection antigen precursor encoded by an alternative open reading frame
CC   (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC   the alternative ORF of macrophage-colony stimulating factor, when
CC   presented by an antigen presenting cell having a human leukocyte antigen
CC   (HLA) class I molecule, effectively induce the activation and
CC   proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC   acids derived from the alternate ORF of macrophage-colony stimulating
CC   factor are useful for enriching selectively a population of T lymphocytes
CC   with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC   characterized by expression of the polypeptide, and for identifying
CC   functional variants and mimetics
XX
SQ   Sequence 15 AA;

  Query Match             89.7%;  Score 61;  DB 3;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.0026;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQE 12
              ||||||||||||
Db          4 PAVVGLSPGEQE 15


RESULT 7
AAB36208
ID   AAB36208 standard; protein; 234 AA.
XX
AC   AAB36208;
XX
DT   15-FEB-2001  (first entry)
XX
DE   Human immune system associated protein HISAP-6.
XX
KW   Human; immune system associated protein; HISAP-6; immune disorder;
KW   infection; autoimmune disease; cancer.
XX
OS   Homo sapiens.
XX
PN   US6135941-A.
XX
PD   24-OCT-2000.
XX
PF   27-MAR-1998;   98US-00049672.
XX
PR   27-MAR-1998;   98US-00049672.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Tang YT,  Yue H,  Lal P,  Corley NC,  Guegler KJ,  Baughn MR;
PI   Hillman JL,  Au-Young J;
XX
DR   WPI; 2001-030926/04.
DR   N-PSDB; AAC66524.
XX
PT   New human immune system associated proteins (HISAP) and polynucleotides
PT   encoding the HISAP, useful for diagnosing, treating or preventing immune
PT   or cell proliferative disorders or infections.
XX
PS   Claim 1; Col 59-60; 54pp; English.
XX
CC   The present invention provides the coding and protein sequences for a
CC   number of human immune system associated proteins (HISAPs). These can be
CC   used in the diagnosis and treatment of various autoimmune disorders,
CC   infections and cell proliferation diseases. The diseases include AIDS,
CC   adult respiratory distress syndrome, anaemia, asthma, atherosclerosis,
CC   Crohn's disease, irritable bowel syndrome, multiple sclerosis, myasthenia
CC   gravis, osteoarthritis, rheumatoid arthritis, scleroderma, systemic lupus
CC   erythematosus, arteriosclerosis, cirrhosis and cancer
XX
SQ   Sequence 234 AA;

  Query Match             66.2%;  Score 45;  DB 4;  Length 234;
  Best Local Similarity   72.7%;  Pred. No. 25;
  Matches    8;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQ 11
              ||:| |||||:
Db         28 PAIVSLSPGER 38


RESULT 8
ABJ10397
ID   ABJ10397 standard; protein; 106 AA.
XX
AC   ABJ10397;
XX
DT   28-NOV-2002  (first entry)
XX
DE   Mutant anti-mesothelin Fv (ST6) variable light chain.
XX
KW   Mutant; mutein; recombinant immunotoxin; reduced liver toxicity; cancer.
XX
OS   Unidentified.
OS   Synthetic.
XX
PN   WO200240545-A2.
XX
PD   23-MAY-2002.
XX
PF   16-NOV-2001; 2001WO-US043602.
XX
PR   17-NOV-2000; 2000US-0249805P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Pastan IH,  Onda M,  Nagata S,  Tsutsumi Y,  Vincent JJ,  Kreitman RJ;
PI   Vasmatzis G,  Lee B;
XX
DR   WPI; 2002-500208/53.
DR   N-PSDB; ABT08085.
XX
PT   Recombinant immunotoxins with reduced liver toxicity for killing
PT   malignant cells bearing antigen comprising mutated framework regions of
PT   antibody heavy and/or light chain or antigen-binding portion of
PT   immunotoxin.
XX
PS   Claim 12; Fig 2; 82pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of
CC   recombinant immunotoxin proteins. The immunotoxin proteins of the
CC   invention contain an antibody (or an antigen-binding fragment) with a
CC   substitution of a negatively charged amino acid for an uncharged or
CC   positively charged amino acid. The immunotoxins of the invention have
CC   reduced liver toxicity. The immunotoxins of the invention are useful for
CC   killing a malignant cell (e.g. a cancer cell). The present amino acid
CC   sequence represents a recombinant immunotoxin of the invention
XX
SQ   Sequence 106 AA;

  Query Match             60.3%;  Score 41;  DB 5;  Length 106;
  Best Local Similarity   63.6%;  Pred. No. 49;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQ 11
              ||::  |||||
Db          8 PAIMSASPGEQ 18


RESULT 9
ABJ10395
ID   ABJ10395 standard; protein; 106 AA.
XX
AC   ABJ10395;
```

```
XX
DT   28-NOV-2002  (first entry)
XX
DE   Mutant anti-Tac Fv (M16) variable light chain.
XX
KW   Mutant; mutein; recombinant immunotoxin; reduced liver toxicity; cancer.
XX
OS   Unidentified.
OS   Synthetic.
XX
PN   WO200240545-A2.
XX
PD   23-MAY-2002.
XX
PF   16-NOV-2001; 2001WO-US043602.
XX
PR   17-NOV-2000; 2000US-0249805P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Pastan IH,  Onda M,  Nagata S,  Tsutsumi Y,  Vincent JJ,  Kreitman RJ;
PI   Vasmatzis G,  Lee B;
XX
DR   WPI; 2002-500208/53.
DR   N-PSDB; ABT08083.
XX
PT   Recombinant immunotoxins with reduced liver toxicity for killing
PT   malignant cells bearing antigen comprising mutated framework regions of
PT   antibody heavy and/or light chain or antigen-binding portion of
PT   immunotoxin.
XX
PS   Claim 8; Fig 2; 82pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of
CC   recombinant immunotoxin proteins. The immunotoxin proteins of the
CC   invention contain an antibody (or an antigen-binding fragment) with a
CC   substitution of a negatively charged amino acid for an uncharged or
CC   positively charged amino acid. The immunotoxins of the invention have
CC   reduced liver toxicity. The immunotoxins of the invention are useful for
CC   killing a malignant cell (e.g. a cancer cell). The present amino acid
CC   sequence represents a recombinant immunotoxin of the invention
XX
SQ   Sequence 106 AA;

  Query Match            60.3%;  Score 41;  DB 5;  Length 106;
  Best Local Similarity  63.6%;  Pred. No. 49;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQ 11
              ||::  |||||
Db          8 PAIMSASPGEQ 18


RESULT 10
AAE10129
ID   AAE10129 standard; protein; 11096 AA.
XX
AC   AAE10129;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin gene, NysC.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; PKS type I.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   Domain          35. .455
FT                   /label= KS3_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          546. .858
FT                   /label= AT3_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          872. .1073
FT                   /label= DH3_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          1381. .1628
FT                   /label= KR3_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          1662. .1735
FT                   /label= ACP3_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          1757. .2180
FT                   /label= KS4_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          2291. .2603
FT                   /label= AT4_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          2617. .2818
FT                   /label= DH4_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          3124. .3371
FT                   /label= KR4_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          3407. .3480
FT                   /label= ACP4_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          3501. .3924
FT                   /label= KS5_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          4032. .4346
FT                   /label= AT5_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          4360. .4561
FT                   /label= DH5_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          4953. .5239
FT                   /label= ER5_domain
FT                   /note= "Enoylreductase (ER) domain"
FT   Domain          5248. .5495
FT                   /label= KR5_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          5528. .5601
FT                   /label= ACP5_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          5623. .6046
FT                   /label= KS6_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          6166. .6478
FT                   /label= AT6_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          6492. .6704
FT                   /label= DH6_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          7038. .7281
FT                   /label= KR6_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          7315. .7388
FT                   /label= ACP6_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          7408. .7831
FT                   /label= KS7_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          7939. .8253
FT                   /label= AT7_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          8267. .8470
FT                   /label= DH7_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          8812. .9086
FT                   /label= KR7_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          9120. .9193
FT                   /label= ACP7_domain
```

```
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          9214. .9637
FT                   /label= KS8_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          9758. .10072
FT                   /label= AT8_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          10086. .10289
FT                   /label= DH8_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          10657. .10904
FT                   /label= KR8_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          10939. .11012
FT                   /label= ACP8_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB000509.
XX
PR   08-FEB-2000; 2000GB-00002840.
PR   10-APR-2000; 2000GB-00008786.
PR   14-APR-2000; 2000GB-00009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
PA   (VALL/) VALLA S.
XX
PI   Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI   Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   N-PSDB; AAD17184.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides, useful
PT   as antibiotics and antifungals.
XX
PS   Claim 15; Page 170-176; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin. The
CC   nystatin PKS is useful as antifungal antibiotics. The present sequence is
CC   a PKS type I encoding Streptomyces noursei nystatin gene, NysC
XX
SQ   Sequence 11096 AA;

  Query Match             60.3%;  Score 41;  DB 4;  Length 11096;
  Best Local Similarity   63.6%;  Pred. No. 8.2e+03;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQ 11
              | | ||:||:|
Db       4998 PEVTGLAPGDQ 5008


RESULT 11
AAB92530
ID   AAB92530 standard; protein; 358 AA.
XX
AC   AAB92530;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:10688.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 10688; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 358 AA;

  Query Match             58.8%;  Score 40;  DB 4;  Length 358;
  Best Local Similarity   58.3%;  Pred. No. 2.7e+02;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 PAVVGLSPGEQE 12
              ||: |  ||:||
Db        213 PAMAGPGPGDQE 224


RESULT 12
ABB97563
ID   ABB97563 standard; protein; 359 AA.
XX
AC   ABB97563;
```

XX
DT 27-JUN-2002 (first entry)
XX
DE Novel human protein SEQ ID NO: 831.
XX
KW Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
KW expressed sequence tag.
XX
OS Homo sapiens.
XX
PN WO200222660-A2.
XX
PD 21-MAR-2002.
XX
PF 10-SEP-2001; 2001WO-US026015.
XX
PR 11-SEP-2000; 2000US-00659671.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR WPI; 2002-292408/33.
DR N-PSDB; ABN32749.
XX
PT An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis.
XX
PS Claim 20; SEQ ID NO 831; 509pp; English.
XX
CC The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a protein of the invention
XX
SQ Sequence 359 AA;

Query Match 58.8%; Score 40; DB 5; Length 359;
Best Local Similarity 58.3%; Pred. No. 2.8e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PAVVGLSPGEQE 12
||: | ||:||
Db 213 PAMAGPGPGDQE 224

RESULT 13
AAB57055
ID AAB57055 standard; protein; 668 AA.
XX
AC AAB57055;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen protein sequence SEQ ID NO:1633.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW vulnerary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease.
XX
OS Homo sapiens.
XX
PN WO200055174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US005988.
XX
PR 12-MAR-1999; 99US-0124270P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587513/55.
DR N-PSDB; AAF16258.
XX
PT Prostate cancer associated gene sequences, referred to as prostate cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as prostate cancer.
XX
PS Claim 11; Page 2091-2094; 2338pp; English.
XX
CC AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioactive, immunomodulatory, muscular, vulnerary, gastrointestinal,
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention
XX
SQ Sequence 668 AA;

Query Match 58.8%; Score 40; DB 3; Length 668;
Best Local Similarity 72.7%; Pred. No. 5.5e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVGLSPGEQEY 13
|||||||: |
Db 27 VVGLSPGQYFY 37

RESULT 14
AAY41875
ID AAY41875 standard; peptide; 18 AA.
XX
AC AAY41875;
XX
DT 09-DEC-1999 (first entry)
XX
DE Rheumatoid arthritis diagnostic protein isoform peptide #26.
XX
KW Human; rheumatoid arthritis; RA; diagnosis; RPI; RADF; detection;
KW rheumatoid arthritis diagnostic feature; ERPI; synovial fluid;
KW rheumatoid arthritis diagnostic protein isoform; screening;
KW expression reference protein isoform; prognosis.
XX
OS Homo sapiens.
XX
PN WO9947925-A2.
XX
PD 23-SEP-1999.
XX
PF 15-MAR-1999; 99WO-GB000763.
XX
PR 13-MAR-1998; 98GB-00005477.

XX
PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI Parekh RB, Patel TP, Townsend RR;
XX
DR WPI; 1999-571871/48.
XX
PT Diagnosis of human rheumatoid arthritis by two-dimensional
PT electrophoresis.
XX
PS Disclosure; Page 18; 157pp; English.
XX
CC A method has been developed for the diagnosis of human rheumatoid
CC arthritis (RA) using two-dimensional electrophoresis to generate a two-
CC dimensional array of features. The method can be used for screening,
CC diagnosis and prognosis of RA in a subject or for monitoring the effect
CC of an anti-RA drug or therapy administered to a subject. The method
CC comprises: (a) analysing a sample of serum or plasma and optionally
CC synovial fluid by two-dimensional electrophoresis, to generate a two-
CC dimensional array of features; (b) identifying at least one chosen
CC feature whose relative abundance correlates with the presence or absence
CC of RA; and (c) comparing the abundance of each chosen feature in the
CC sample with the abundance of that chosen feature in serum or plasma from
CC one or more persons without RA, where the relative abundance of the
CC chosen feature or features in the sample indicates the presence or
CC absence of RA in the subject. The method can also be used in clinical
CC studies for testing drugs for therapy of RA, for purification of RA-
CC diagnostic protein isoforms (RPIs), and for production of antibodies to
CC RPIs. The RA-diagnostic feature (RADF) proteins can be used to identify
CC compounds that promote or inhibit their activity, which are then used as
CC RA drugs. Nucleic acid encoding RADFs can be used in gene therapy
CC protocols. AAY41844 to AAY42100 represent RPI peptides, AAY42101 to
CC AAY42103 represent expression reference protein isoform peptides and
CC AAZ25066 to AAZ25068 represent degenerate probes for RPIs, which are all
CC used in the exemplification of the present invention
XX
SQ Sequence 18 AA;

Query Match 57.4%; Score 39; DB 2; Length 18;
Best Local Similarity 63.6%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PAVVGLSPGEQ 11
|| : |||||:
Db 8 PATLSLSPGER 18

RESULT 15
AAU25388
ID AAU25388 standard; peptide; 18 AA.
XX
AC AAU25388;
XX
DT 18-DEC-2001 (first entry)
XX
DE Schizophrenia-Associated Protein Isoform (SPI) peptide #617.
XX
KW Schizophrenia-associated protein isoform; SPI; SPI-206; SPI-238; SPI-240;
KW neuroleptic; gene therapy; cerebrospinal fluid; serum; plasma.
XX
OS Homo sapiens.
XX
PN WO200162785-A2.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-GB000792.
XX
PR 24-FEB-2000; 2000GB-00004415.
PR 28-DEC-2000; 2000US-00750395.
XX
PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI Herath HMAC, Parekh RB, Rohlff C, Terrett JA, Tyson KL;
XX
DR WPI; 2001-570624/64.
XX
PT New schizophrenia associated protein isoforms and encoding nucleic acid
PT molecules, useful for treatment, diagnosis and prognosis of schizophrenia
PT and screening for potential drugs for treatment and new drug targets.
XX
PS Disclosure; Page 41; 148pp; English.
XX
CC The sequence represents a schizophrenia-associated protein isoform (SPI).
CC These protein isoforms, e.g. SPI-206, SPI-238 and SPI-240 are detectable
CC in cerebrospinal fluid, serum or plasma and are useful markers of
CC schizophrenia. The sequences can be used for treatment and diagnosis of
CC schizophrenia, screening, prognosis, monitoring the results of therapy,
CC identifying patients most likely to respond to a particular therapy and
CC identification of new targets for drug treatment. SPI DNA is useful as a
CC nucleic acid probe to detect the presence of nucleic acids or SPIs
XX
SQ Sequence 18 AA;

Query Match 57.4%; Score 39; DB 4; Length 18;
Best Local Similarity 63.6%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PAVVGLSPGEQ 11
|| : |||||:
Db 8 PATLSLSPGER 18

Search completed: May 7, 2004, 12:33:46
Job time : 37.27 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:26:40 ; Search time 24.31 Seconds
(without alignments)
168.726 Million cell updates/sec

Title: US-09-786-214A-14
Perfect score: 65
Sequence: 1 LPAVVGLSPGEQE 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	49	75.4	41	11	Q8K408	Q8k408 rattus norv
2	44	67.7	381	16	Q9RRL5	Q9rrl5 deinococcus
3	43.5	66.9	614	16	Q8Y280	Q8y280 ralstonia s
4	43	66.2	821	17	Q9HPR8	Q9hpr8 halobacteri
5	42	64.6	1541	6	Q8HXL3	Q8hxl3 sus scrofa
6	41	63.1	326	1	Q9UXP0	Q9uxp0 methanolobu
7	41	63.1	579	4	Q8N158	Q8n158 homo sapien
8	41	63.1	11096	2	Q9L4W3	Q914w3 streptomyce
9	40	61.5	154	17	Q8ZYJ3	Q8zyj3 pyrobaculum
10	40	61.5	242	16	Q7WCX1	Q7wcx1 bordetella
11	40	61.5	261	16	Q8NQA2	Q8nqa2 corynebacte
12	40	61.5	298	16	Q7W5D6	Q7w5d6 bordetella
13	40	61.5	298	16	Q7VTI2	Q7vti2 bordetella
14	40	61.5	355	10	Q947A7	Q947a7 nitellopsis
15	40	61.5	357	4	Q8NHX1	Q8nhx1 homo sapien
16	40	61.5	358	4	Q9NWD0	Q9nwd0 homo sapien
17	40	61.5	386	4	Q9NWM3	Q9nwm3 homo sapien
18	40	61.5	428	16	Q92AG9	Q92ag9 listeria in
19	40	61.5	428	16	Q8Y661	Q8y661 listeria mo
20	40	61.5	540	16	Q9RR71	Q9rr71 deinococcus
21	40	61.5	753	16	Q89T31	Q89t31 bradyrhizob
22	39	60.0	197	17	Q8PWI9	Q8pwi9 methanosarc
23	39	60.0	209	17	Q8TTT1	Q8ttt1 methanosarc
24	39	60.0	209	17	Q8TH81	Q8th81 methanosarc
25	39	60.0	319	16	Q8ZKQ5	Q8zkq5 salmonella
26	39	60.0	319	16	Q8Z2X4	Q8z2x4 salmonella
27	39	60.0	407	2	Q9LCW0	Q9lcw0 streptomyce
28	39	60.0	626	16	Q7VWC3	Q7vwc3 bordetella
29	39	60.0	627	16	Q7WH05	Q7wh05 bordetella
30	39	60.0	627	16	Q7W9Q3	Q7w9q3 bordetella
31	39	60.0	1400	11	Q7TME1	Q7tme1 mus musculu
32	38.5	59.2	656	16	Q9RW14	Q9rw14 deinococcus
33	38	58.5	62	16	Q7VER3	Q7ver3 mycobacteri
34	38	58.5	136	11	Q61061	Q61061 mus musculu
35	38	58.5	152	17	Q9V2S2	Q9v2s2 pyrococcus
36	38	58.5	152	17	Q8TZK1	Q8tzk1 pyrococcus
37	38	58.5	155	17	O57778	O57778 pyrococcus
38	38	58.5	177	16	Q8U7R5	Q8u7r5 agrobacteri
39	38	58.5	212	5	Q9NA60	Q9na60 caenorhabdi
40	38	58.5	222	17	Q97WC8	Q97wc8 sulfolobus
41	38	58.5	313	5	Q7YTA2	Q7yta2 glomeris ma
42	38	58.5	325	16	Q9CBS2	Q9cbs2 mycobacteri
43	38	58.5	335	16	Q82RR7	Q82rr7 streptomyce
44	38	58.5	352	16	Q8G7R2	Q8g7r2 bifidobacte
45	38	58.5	383	16	Q7V8U3	Q7v8u3 prochloroco

ALIGNMENTS

RESULT 1
Q8K408
ID Q8K408 PRELIMINARY; PRT; 41 AA.
AC Q8K408;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Truncated macrophage colony stimulating factor.
GN CSF1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEW.tl;
RX MEDLINE=22069908; PubMed=12074592;
RA Dobbins D.E., Sood R., Hashiramoto A., Hansen C.T., Wilder R.L.,
RA Remmers E.F.;
RT "Mutation of macrophage colony stimulating factor (Csf1) causes
RT osteopetrosis in the tl rat.";
RL Biochem. Biophys. Res. Commun. 294:1114-1120(2002).
DR EMBL; AF514357; AAM54137.1; -.
SQ SEQUENCE 41 AA; 4178 MW; 1D342C19BD18AA41 CRC64;

Query Match 75.4%; Score 49; DB 11; Length 41;
Best Local Similarity 76.9%; Pred. No. 0.44;
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 LPAVVGLSPGEQE 13
              ||| |||| |||
Db         21 LPAAAGLSPREQE 33
```

RESULT 2
Q9RRL5
ID Q9RRL5 PRELIMINARY; PRT; 381 AA.
AC Q9RRL5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

```
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cytochrome P450.
GN   DR2473.
OS   Deinococcus radiodurans.
OC   Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AE002076; AAF12016.1; -.
DR   PIR; F75270; F75270.
DR   TIGR; DR2473; -.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase; Complete proteome.
SQ   SEQUENCE   381 AA;  41940 MW;  F191EA69F1797B53 CRC64;

  Query Match             67.7%;  Score 44;  DB 16;  Length 381;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSP 9
              |||||||||
Db         52 LPAVVGLSP 60


RESULT 3
Q8Y280
ID   Q8Y280      PRELIMINARY;      PRT;   614 AA.
AC   Q8Y280;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Probable ATP-binding transport ABC transporter protein.
GN   RSC0456 OR RS04444.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Burkholderiaceae; Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646059; CAD13984.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR   GO; GO:0000166; F:nucleotide binding; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR001623; DnaJ_N.
DR   Pfam; PF00005; ABC_tran; 1.
DR   ProDom; PD000006; ABC_transporter; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR   PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
DR   PROSITE; PS00636; DNAJ_1; 1.
KW   Complete proteome.
SQ   SEQUENCE   614 AA;  69240 MW;  E293355B85872142 CRC64;

  Query Match             66.9%;  Score 43.5;  DB 16;  Length 614;
  Best Local Similarity   45.8%;  Pred. No. 63;
  Matches   11;  Conservative    2;  Mismatches    0;  Indels   11;  Gaps    1;

Qy          1 LPAVVG----------LSPGEQE 13
              |||:||           ||||||:
Db        501 LPALVGLLDEVSNWSLRLSPGEQQ 524


RESULT 4
Q9HPR8
ID   Q9HPR8      PRELIMINARY;      PRT;   821 AA.
AC   Q9HPR8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   DNA helicase.
GN   HEL OR VNG1501G.
OS   Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR   EMBL; AE005064; AAG19799.1; -.
DR   PIR; C84304; C84304.
DR   GO; GO:0004386; F:helicase activity; IEA.
DR   GO; GO:0000166; F:nucleotide binding; IEA.
DR   GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001478; PDZ.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00228; PDZ; 1.
KW   Helicase; Complete proteome.
SQ   SEQUENCE   821 AA;  89848 MW;  C454C76B984A5702 CRC64;

  Query Match             66.2%;  Score 43;  DB 17;  Length 821;
  Best Local Similarity   90.0%;  Pred. No. 1e+02;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AVVGLSPGEQ 12
              |||||| ||
Db        326 AVVGLSPAEQ 335


RESULT 5
Q8HXL3
ID   Q8HXL3      PRELIMINARY;      PRT;  1541 AA.
```

```
AC   Q8HXL3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21845885; PubMed=11856879;
RA   Drogemuller C., Kuiper H., Voss-Nemitz R., Brenig B., Distl O.,
RA   Leeb T.;
RT   "Molecular characterization and chromosome assignment of the porcine
RT   gene COX7A1 coding for the muscle specific cytochrome c oxidase
RT   subunit VIIa-M.";
RL   Cytogenet. Cell Genet. 94:190-193(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Leeb T.;
RL   Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ410870; CAD56046.1; -.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00400; WD40; 12.
DR   PROSITE; PS50082; WD_REPEATS_2; 1.
DR   PROSITE; PS50294; WD_REPEATS_REGION; 3.
KW   Hypothetical protein.
SQ   SEQUENCE   1541 AA;  168898 MW;  81B8882854AF4F1E CRC64;

  Query Match             64.6%;  Score 42;  DB 6;  Length 1541;
  Best Local Similarity   58.3%;  Pred. No. 2.9e+02;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQ 12
              :|: : ||||||
Db        786 MPSEISLSPGEQ 797


RESULT 6
Q9UXP0
ID   Q9UXP0      PRELIMINARY;      PRT;   326 AA.
AC   Q9UXP0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   F420-dependent N5,N10-methylene-tetrahydromethanopterin reductase,
DE   putative.
GN   FFDA.
OS   Methanolobus tindarius.
OC   Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC   Methanosarcinales; Methanosarcinaceae; Methanolobus.
OX   NCBI_TaxID=2221;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 2278;
RA   Westenberg D.J., Braune A., Ruppert C., Mueller V., Herzberg C.,
RA   Gottschalk G., Blaut M.;
RT   "The F420H2-dehydrogenase from Methanolobus tindarius: Cloning of the
RT   ffd operon and expression of the genes in Escherichia coli.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ011519; CAB56639.1; -.
DR   PIR; T45226; T45226.
DR   InterPro; IPR002103; Bac_luciferase.
DR   Pfam; PF00296; bac_luciferase; 1.
SQ   SEQUENCE   326 AA;  34043 MW;  16F3AB9733A45D82 CRC64;

  Query Match             63.1%;  Score 41;  DB 1;  Length 326;
  Best Local Similarity   70.0%;  Pred. No. 88;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AVVGLSPGEQ 12
              |::|| ||||
Db         88 AILGLGPGEQ 97


RESULT 7
Q8N158
ID   Q8N158      PRELIMINARY;      PRT;   579 AA.
AC   Q8N158;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Similar to cerebroglycan (Hypothetical protein FLJ38962).
GN   DKFZP547M109.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA   Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA   Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA   Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Bloecker H., Boecher M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC027972; AAH27972.1; -.
DR   EMBL; AK096281; BAC04745.1; -.
DR   EMBL; AL834418; CAD39080.1; -.
DR   Genew; HGNC:4450; GPC2.
DR   GO; GO:0005578; C:extracellular matrix; IEA.
DR   InterPro; IPR001863; Glypican.
DR   Pfam; PF01153; Glypican; 1.
DR   PROSITE; PS01207; GLYPICAN; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   579 AA;  62829 MW;  1630E3A22BB83DFA CRC64;

  Query Match             63.1%;  Score 41;  DB 4;  Length 579;
  Best Local Similarity   75.0%;  Pred. No. 1.6e+02;
  Matches    9;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQ 12
              || ||| || ||
Db        439 LPPVVGGSPAEQ 450


RESULT 8
Q9L4W3
ID   Q9L4W3      PRELIMINARY;      PRT; 11096 AA.
AC   Q9L4W3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NysC.
GN   NYSC.
OS   Streptomyces noursei.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=ATCC 11455;
RX   MEDLINE=20334850; PubMed=10873841;
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT   Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT   deduction of the biosynthetic pathway.";
RL   Chem. Biol. 7:395-403(2000).
DR   EMBL; AF263912; AAF71776.1; -.
DR   HSSP; P25715; 1MLA.
DR   GO; GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0004314; F:[acyl-carrier protein] S-malonyltransferase. . .; IEA.
DR   GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR001227; Ac_trans.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl_synth.
DR   InterPro; IPR006162; Ppantne_S.
DR   InterPro; IPR006163; Pp_bind.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00107; ADH_zinc_N; 1.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   11096 AA;  1150415 MW;  776CAEAFCAE551DD CRC64;

  Query Match           63.1%;  Score 41;  DB 2;  Length 11096;
  Best Local Similarity  63.6%;  Pred. No. 3.3e+03;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 PAVVGLSPGEQ 12
              | | ||:||:|
Db       4998 PEVTGLAPGDQ 5008


RESULT 9
Q8ZYJ3
ID   Q8ZYJ3      PRELIMINARY;      PRT;   154 AA.
AC   Q8ZYJ3;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein PAE0746.
GN   PAE0746.
OS   Pyrobaculum aerophilum.
OC   Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC   Thermoproteaceae; Pyrobaculum.
OX   NCBI_TaxID=13773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX   MEDLINE=21664397; PubMed=11792869;
RA   Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA   Miller J.H.;
RT   "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT   aerophilum.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR   EMBL; AE009783; AAL63000.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   154 AA;  17403 MW;  1C3D8BCB40324766 CRC64;

  Query Match           61.5%;  Score 40;  DB 17;  Length 154;
  Best Local Similarity  66.7%;  Pred. No. 60;
  Matches   10;  Conservative    1;  Mismatches    2;  Indels    2;  Gaps    1;

Qy          1 LPAVVGLS--PGEQE 13
              ||  ||:|  |||||
Db         32 LPPDVGISYTPGEQE 46


RESULT 10
Q7WCX1
ID   Q7WCX1      PRELIMINARY;      PRT;   242 AA.
AC   Q7WCX1;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Bacteriophage-related DNA polymerase.
GN   BB3808.
OS   Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Bordetella.
OX   NCBI_TaxID=518;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RB50 / ATCC BAA-588;
RX   MEDLINE=22827954; PubMed=12910271;
RA   Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA   Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA   Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA   Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA   Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA   Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA   Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA   Rabbinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA   Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA   Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT   "Comparative analysis of the genome sequences of Bordetella pertussis,
RT   Bordetella parapertussis and Bordetella bronchiseptica.";
RL   Nat. Genet. 35:32-40(2003).
DR   EMBL; BX640448; CAE35782.1; -.
KW   Complete proteome.
SQ   SEQUENCE   242 AA;  25627 MW;  EECAD9B319823284 CRC64;

  Query Match           61.5%;  Score 40;  DB 16;  Length 242;
  Best Local Similarity  80.0%;  Pred. No. 96;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 VVGLSPGEQE 13
              ||| :|||||
Db         96 VVGEAPGEQE 105


RESULT 11
Q8NQA2
ID   Q8NQA2      PRELIMINARY;      PRT;   261 AA.
AC   Q8NQA2;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein Cgl1536.
GN   CGL1536.
OS   Corynebacterium glutamicum (Brevibacterium flavum).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX   NCBI_TaxID=1718;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA   Nakagawa S.;
RT   "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP005278; BAB98929.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   261 AA;  27957 MW;  4D8D51D4DCA3A210 CRC64;

  Query Match           61.5%;  Score 40;  DB 16;  Length 261;
  Best Local Similarity  53.8%;  Pred. No. 1e+02;
```

```
  Matches    7;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQE 13
              |||:  :|||| :
Db        137 LPAIATVSPGEAD 149


RESULT 12
Q7W5D6
ID   Q7W5D6      PRELIMINARY;      PRT;   298 AA.
AC   Q7W5D6;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Bacteriophage-related DNA polymerase.
GN   BPP3357.
OS   Bordetella parapertussis.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Bordetella.
OX   NCBI_TaxID=519;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=12822 / ATCC BAA-587;
RX   MEDLINE=22827954; PubMed=12910271;
RA   Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA   Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA   Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA   Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA   Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA   Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA   Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA   Rabbinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA   Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA   Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT   "Comparative analysis of the genome sequences of Bordetella pertussis,
RT   Bordetella parapertussis and Bordetella bronchiseptica.";
RL   Nat. Genet. 35:32-40(2003).
DR   EMBL; BX640433; CAE38642.1; -.
KW   Complete proteome.
SQ   SEQUENCE   298 AA;  31202 MW;  A858C8199E2AC8B7 CRC64;

  Query Match             61.5%;  Score 40;  DB 16;  Length 298;
  Best Local Similarity   80.0%;  Pred. No. 1.2e+02;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 VVGLSPGEQE 13
              ||| :|||||
Db        152 VVGEAPGEQE 161


RESULT 13
Q7VTI2
ID   Q7VTI2      PRELIMINARY;      PRT;   298 AA.
AC   Q7VTI2;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Bacteriophage-related DNA polymerase.
GN   BP3556.
OS   Bordetella pertussis.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Bordetella.
OX   NCBI_TaxID=520;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RX   MEDLINE=22827954; PubMed=12910271;
RA   Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA   Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA   Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA   Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA   Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA   Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA   Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA   Rabbinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA   Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA   Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT   "Comparative analysis of the genome sequences of Bordetella pertussis,
RT   Bordetella parapertussis and Bordetella bronchiseptica.";
RL   Nat. Genet. 35:32-40(2003).
DR   EMBL; BX640421; CAE43815.1; -.
KW   Complete proteome.
SQ   SEQUENCE   298 AA;  31231 MW;  9E4BC84B419506F3 CRC64;

  Query Match             61.5%;  Score 40;  DB 16;  Length 298;
  Best Local Similarity   80.0%;  Pred. No. 1.2e+02;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 VVGLSPGEQE 13
              ||| :|||||
Db        152 VVGEAPGEQE 161


RESULT 14
Q947A7
ID   Q947A7      PRELIMINARY;      PRT;   355 AA.
AC   Q947A7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Enolase (EC 4.2.1.11) (2-phosphoglycerate dehydratase) (2-phospho-D-
DE   glycerate hydro-lyase) (Fragment).
GN   ENO.
OS   Nitellopsis obtusa.
OC   Eukaryota; Viridiplantae; Streptophyta; Charales; Characeae;
OC   Nitellopsis.
OX   NCBI_TaxID=40811;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=X-854;
RX   MEDLINE=21437986; PubMed=11526220;
RA   Keeling P.J., Palmer J.D.;
RT   "Lateral transfer at the gene and subgenic levels in the evolution of
RT   eukaryotic enolase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:10745-10750(2001).
CC   -!- CATALYTIC ACTIVITY: 2-PHOSPHO-D-GLYCERATE = PHOSPHOENOLPYRUVATE +
CC       H(2)O.
CC   -!- COFACTOR: MAGNESIUM IS REQUIRED FOR CATALYSIS AND FOR STABILIZING
CC       THE DIMER (BY SIMILARITY).
CC   -!- PATHWAY: GLYCOLYSIS.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ENOLASE FAMILY.
DR   EMBL; AF348916; AAL05455.1; -.
DR   GO; GO:0000015; C:phosphopyruvate hydratase complex; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0000287; F:magnesium ion binding; IEA.
DR   GO; GO:0004634; F:phosphopyruvate hydratase activity; IEA.
DR   GO; GO:0006096; P:glycolysis; IEA.
DR   InterPro; IPR000941; Enolase.
DR   Pfam; PF00113; enolase; 1.
DR   Pfam; PF03952; enolase_N; 1.
DR   PRINTS; PR00148; ENOLASE.
DR   ProDom; PD000902; Enolase; 1.
DR   PROSITE; PS00164; ENOLASE; 1.
KW   Glycolysis; Lyase; Magnesium.
FT   NON_TER       1      1
FT   NON_TER     355    355
SQ   SEQUENCE   355 AA;  38295 MW;  400DF160087DE450 CRC64;

  Query Match             61.5%;  Score 40;  DB 10;  Length 355;
  Best Local Similarity   54.5%;  Pred. No. 1.4e+02;
  Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 PAVVGLSPGEQ 12
              |||:|: | :|
```

```
Db         28 PAVIGMDPADQ 38


RESULT 15
Q8NHX1
ID   Q8NHX1      PRELIMINARY;      PRT;   357 AA.
AC   Q8NHX1;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Extracellular signal-related kinase 1b.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Aebersold D.M., Yung Y., Seger R.;
RT   "Properties of human ERK1b.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY033607; AAK52329.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004707; F:MAP kinase activity; IEA.
DR   GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR   GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   InterPro; IPR008349; Erk_1_2_MAPK.
DR   InterPro; IPR008350; Erk_3_4_MAPK.
DR   InterPro; IPR003527; MAP_kin.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR01770; ERK1ERK2MAPK.
DR   PRINTS; PR01771; ERK3ERK4MAPK.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS01351; MAPK; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   ATP-binding; Kinase; Transferase.
SQ   SEQUENCE   357 AA;  40062 MW;  58C92773983ADA79 CRC64;

  Query Match             61.5%;  Score 40;  DB 4;  Length 357;
  Best Local Similarity   72.7%;  Pred. No. 1.4e+02;
  Matches    8;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 PAVVGLSPGEQ 12
              || |||  |||
Db        344 PAAVGLGAGEQ 354


Search completed: May 7, 2004, 12:37:56
Job time : 25.4767 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:19:40 ; Search time 36.27 Seconds
(without alignments)
101.272 Million cell updates/sec

Title: US-09-786-214A-14
Perfect score: 65
Sequence: 1 LPAVVGLSPGEQE 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	65	100.0	13	3	AAY84268	Aay84268 Peptide d
2	65	100.0	14	3	AAY84266	Aay84266 Peptide d
3	65	100.0	15	3	AAY84269	Aay84269 Peptide d
4	65	100.0	20	3	AAY84265	Aay84265 Truncated
5	65	100.0	25	3	AAY84264	Aay84264 Peptide o
6	61	93.8	13	3	AAY84267	Aay84267 Peptide d
7	45	69.2	234	4	AAB36208	Aab36208 Human imm
8	42	64.6	470	6	AAE34724	Aae34724 Streptomy
9	42	64.6	475	6	AAE34732	Aae34732 Streptomy
10	42	64.6	475	6	AAE34729	Aae34729 Streptomy
11	41	63.1	106	5	ABJ10397	Abj10397 Mutant an
12	41	63.1	106	5	ABJ10395	Abj10395 Mutant an
13	41	63.1	530	7	ADD49105	Add49105 Human NOV
14	41	63.1	549	7	ADD49091	Add49091 Human NOV
15	41	63.1	579	5	ABG70277	Abg70277 Human Gly
16	41	63.1	579	5	ABG97356	Abg97356 Human CGD
17	41	63.1	579	6	ABR39111	Abr39111 Human GPC
18	41	63.1	579	7	ADD49087	Add49087 Human NOV
19	41	63.1	579	7	ADD49107	Add49107 Human NOV
20	41	63.1	579	7	ADD49089	Add49089 Human NOV
21	41	63.1	592	7	ADD49099	Add49099 Human NOV
22	41	63.1	11096	4	AAE10129	Aae10129 Streptomy
23	40	61.5	261	4	AAG91441	Aag91441 C glutami
24	40	61.5	306	4	ABG24698	Abg24698 Novel hum
25	40	61.5	358	4	AAB92530	Aab92530 Human pro
26	40	61.5	359	5	ABB97563	Abb97563 Novel hum
27	40	61.5	428	5	ABB47705	Abb47705 Listeria
28	39	60.0	18	2	AAY41875	Aay41875 Rheumatoi
29	39	60.0	18	4	AAU25388	Aau25388 Schizophr
30	39	60.0	18	4	AAU15732	Aau15732 Schizophr
31	39	60.0	18	5	ABG78871	Abg78871 Multiple
32	39	60.0	23	5	ABP62618	Abp62618 Human imm
33	39	60.0	74	6	ABJ18694	Abj18694 Antibody
34	39	60.0	88	3	AAY56655	Aay56655 Partial p
35	39	60.0	93	5	AAU80983	Aau80983 Human ant
36	39	60.0	94	7	ADD69248	Add69248 Human lig
37	39	60.0	95	6	ABO27154	Abo27154 Human ger
38	39	60.0	95	6	ABO27153	Abo27153 Human ger
39	39	60.0	96	6	ABO27155	Abo27155 Human ger
40	39	60.0	96	6	ABO27150	Abo27150 Human ger
41	39	60.0	100	5	AAE23987	Aae23987 Human MOG
42	39	60.0	104	2	AAW31723	Aaw31723 Alpha lig
43	39	60.0	104	2	AAW26795	Aaw26795 Anti-gp54
44	39	60.0	105	7	ABR61567	Abr61567 HIV-1 neu
45	39	60.0	106	2	AAR41234	Aar41234 Monoclona

ALIGNMENTS

RESULT 1
AAY84268
ID AAY84268 standard; peptide; 13 AA.
XX
AC AAY84268;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Example 2; Page 40; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying

```
CC   functional variants and mimetics
XX
SQ   Sequence 13 AA;

  Query Match             100.0%;  Score 65;  DB 3;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 0.00057;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQE 13
              |||||||||||||
Db          1 LPAVVGLSPGEQE 13


RESULT 2
AAY84266
ID   AAY84266 standard; peptide; 14 AA.
XX
AC   AAY84266;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW   tumour rejection antigen; macrophage colony stimulating gene;
KW   macrophage-colony stimulating factor; antigen presenting cell;
KW   human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO200013699-A1.
XX
PD   16-MAR-2000.
XX
PF   03-SEP-1999;   99WO-US020344.
XX
PR   04-SEP-1998;   98US-0099077P.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Probst-Kepper M,  Van Den Eynde B,  Boon-Falleur T;
XX
DR   WPI; 2000-256859/22.
XX
PT   Isolated polypeptide used to treat subjects having a disorder
PT   characterized by expression of alternative open reading frame macrophage-
PT   colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS   Claim 2; Page 39; 74pp; English.
XX
CC   The present sequence represents a peptide which is derived from a tumour
CC   rejection antigen precursor encoded by an alternative open reading frame
CC   (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC   the alternative ORF of macrophage-colony stimulating factor, when
CC   presented by an antigen presenting cell having a human leukocyte antigen
CC   (HLA) class I molecule, effectively induce the activation and
CC   proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC   acids derived from the alternate ORF of macrophage-colony stimulating
CC   factor are useful for enriching selectively a population of T lymphocytes
CC   with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC   characterized by expression of the polypeptide, and for identifying
CC   functional variants and mimetics
XX
SQ   Sequence 14 AA;

  Query Match             100.0%;  Score 65;  DB 3;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.00061;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQE 13
              |||||||||||||
Db          1 LPAVVGLSPGEQE 13


RESULT 3
AAY84269
ID   AAY84269 standard; peptide; 15 AA.
XX
AC   AAY84269;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW   tumour rejection antigen; macrophage colony stimulating gene;
KW   macrophage-colony stimulating factor; antigen presenting cell;
KW   human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO200013699-A1.
XX
PD   16-MAR-2000.
XX
PF   03-SEP-1999;   99WO-US020344.
XX
PR   04-SEP-1998;   98US-0099077P.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Probst-Kepper M,  Van Den Eynde B,  Boon-Falleur T;
XX
DR   WPI; 2000-256859/22.
XX
PT   Isolated polypeptide used to treat subjects having a disorder
PT   characterized by expression of alternative open reading frame macrophage-
PT   colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS   Example 2; Page 40; 74pp; English.
XX
CC   The present sequence represents a peptide which is derived from a tumour
CC   rejection antigen precursor encoded by an alternative open reading frame
CC   (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC   the alternative ORF of macrophage-colony stimulating factor, when
CC   presented by an antigen presenting cell having a human leukocyte antigen
CC   (HLA) class I molecule, effectively induce the activation and
CC   proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC   acids derived from the alternate ORF of macrophage-colony stimulating
CC   factor are useful for enriching selectively a population of T lymphocytes
CC   with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC   characterized by expression of the polypeptide, and for identifying
CC   functional variants and mimetics
XX
SQ   Sequence 15 AA;

  Query Match             100.0%;  Score 65;  DB 3;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.00066;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQE 13
              |||||||||||||
Db          3 LPAVVGLSPGEQE 15


RESULT 4
AAY84265
ID   AAY84265 standard; peptide; 20 AA.
XX
AC   AAY84265;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Truncated macrophage colony stimulating factor tumour antigen.
XX
KW   tumour rejection antigen; macrophage colony stimulating gene;
```

KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
DR N-PSDB; AAZ99675.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Claim 3; Page 64; 74pp; English.
XX
CC The present sequence represents a truncated tumour rejection antigen
CC precursor, and is encoded by a truncated alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics
XX
SQ Sequence 20 AA;

Query Match 100.0%; Score 65; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.0009;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPGEQE 13
||||||||||||
Db 4 LPAVVGLSPGEQE 16

RESULT 5
AAY84264
ID AAY84264 standard; peptide; 25 AA.
XX
AC AAY84264;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide of alternate reading frame of macrophage colony stimulating gene.
XX
KW Renal cell carcinoma; antigen; cytotoxic T lymphocyte;
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
DR N-PSDB; AAZ99672.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Claim 1; Page 64; 74pp; English.
XX
CC The present sequence represents a tumour rejection antigen precursor, and
CC is encoded by an alternative open reading frame (ORF) of human macrophage
CC colony stimulating gene. Peptides derived from the alternative ORF of
CC macrophage-colony stimulating factor, when presented by an antigen
CC presenting cell having a human leukocyte antigen (HLA) class I molecule,
CC effectively induce the activation and proliferation of CD8+ cytotoxic T
CC lymphocytes. Polypeptide and nucleic acids derived from the alternate ORF
CC of macrophage-colony stimulating factor are useful for enriching
CC selectively a population of T lymphocytes with CD8+ T lymphocytes. They
CC are also useful for diagnosing a disorder characterized by expression of
CC the polypeptide, and for identifying functional variants and mimetics
XX
SQ Sequence 25 AA;

Query Match 100.0%; Score 65; DB 3; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPGEQE 13
||||||||||||
Db 4 LPAVVGLSPGEQE 16

RESULT 6
AAY84267
ID AAY84267 standard; peptide; 13 AA.
XX
AC AAY84267;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX

PS Example 2; Page 40; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics
XX
SQ Sequence 13 AA;

Query Match 93.8%; Score 61; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.0026;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQE 13
||||||||||||
Db 1 PAVVGLSPGEQE 12

RESULT 7
AAB36208
ID AAB36208 standard; protein; 234 AA.
XX
AC AAB36208;
XX
DT 15-FEB-2001 (first entry)
XX
DE Human immune system associated protein HISAP-6.
XX
KW Human; immune system associated protein; HISAP-6; immune disorder;
KW infection; autoimmune disease; cancer.
XX
OS Homo sapiens.
XX
PN US6135941-A.
XX
PD 24-OCT-2000.
XX
PF 27-MAR-1998; 98US-00049672.
XX
PR 27-MAR-1998; 98US-00049672.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Yue H, Lal P, Corley NC, Guegler KJ, Baughn MR;
PI Hillman JL, Au-Young J;
XX
DR WPI; 2001-030926/04.
DR N-PSDB; AAC66524.
XX
PT New human immune system associated proteins (HISAP) and polynucleotides
PT encoding the HISAP, useful for diagnosing, treating or preventing immune
PT or cell proliferative disorders or infections.
XX
PS Claim 1; Col 59-60; 54pp; English.
XX
CC The present invention provides the coding and protein sequences for a
CC number of human immune system associated proteins (HISAPs). These can be
CC used in the diagnosis and treatment of various autoimmune disorders,
CC infections and cell proliferation diseases. The diseases include AIDS,
CC adult respiratory distress syndrome, anaemia, asthma, atherosclerosis,
CC Crohn's disease, irritable bowel syndrome, multiple sclerosis, myasthenia
CC gravis, osteoarthritis, rheumatoid arthritis, scleroderma, systemic lupus
CC erythematosus, arteriosclerosis, cirrhosis and cancer
XX
SQ Sequence 234 AA;

Query Match 69.2%; Score 45; DB 4; Length 234;
Best Local Similarity 72.7%; Pred. No. 26;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQ 12
||:| |||||:
Db 28 PAIVSLSPGER 38

RESULT 8
AAE34724
ID AAE34724 standard; protein; 470 AA.
XX
AC AAE34724;
XX
DT 14-MAY-2003 (first entry)
XX
DE Streptomyces rimosus ema3 protein.
XX
KW P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;
KW emamectin; insecticide; ema3 protein.
XX
OS Streptomyces rimosus.
XX
PN WO200292801-A2.
XX
PD 21-NOV-2002.
XX
PF 15-MAY-2002; 2002WO-EP005363.
XX
PR 16-MAY-2001; 2001US-0291149P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
PI Buckel TG;
XX
DR WPI; 2003-140280/13.
DR N-PSDB; AAD53019.
XX
PT Novel polypeptide that exhibits enzymatic activity of P450 monooxygenase
PT and capable of regioselectively oxidizing alcohol of avermectin useful
PT for making emamectin from avermectin.
XX
PS Claim 17; Page 107-108; 157pp; English.
XX
CC The present invention relates to novel proteins that exhibit an enzymatic
CC activity of P450 monooxygenase and capable of regioselectively oxidising
CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
CC The invention also relates to ferrodoxins and ferrodoxin reductases that
CC are active with the P450 monooxygenases. Sequences of the invention are
CC useful for producing 4-keto-avermectin from avermectin, which is useful
CC for producing emamectin. Emamectin is useful as an insecticide. The
CC present sequence is Streptomyces rimosus ema3 protein
XX
SQ Sequence 470 AA;

Query Match 64.6%; Score 42; DB 6; Length 470;
Best Local Similarity 72.7%; Pred. No. 1.8e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPGE 11
||: ||| |||
Db 44 LPSYVGLHPGE 54

RESULT 9
AAE34732
ID AAE34732 standard; protein; 475 AA.
XX
AC AAE34732;
XX

DT 14-MAY-2003 (first entry)
XX
DE Streptomyces rimosus ema11 protein.
XX
KW P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;
KW emamectin; insecticide; ema11 protein.
XX
OS Streptomyces rimosus.
XX
PN WO200292801-A2.
XX
PD 21-NOV-2002.
XX
PF 15-MAY-2002; 2002WO-EP005363.
XX
PR 16-MAY-2001; 2001US-0291149P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
PI Buckel TG;
XX
DR WPI; 2003-140280/13.
DR N-PSDB; AAD53027.
XX
PT Novel polypeptide that exhibits enzymatic activity of P450 monooxygenase
PT and capable of regioselectively oxidizing alcohol of avermectin useful
PT for making emamectin from avermectin.
XX
PS Claim 17; Page 121-122; 157pp; English.
XX
CC The present invention relates to novel proteins that exhibit an enzymatic
CC activity of P450 monooxygenase and capable of regioselectively oxidising
CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
CC The invention also relates to ferrodoxins and ferrodoxin reductases that
CC are active with the P450 monooxygenases. Sequences of the invention are
CC useful for producing 4-keto-avermectin from avermectin, which is useful
CC for producing emamectin. Emamectin is useful as an insecticide. The
CC present sequence is Streptomyces rimosus ema11 protein
XX
SQ Sequence 475 AA;

Query Match 64.6%; Score 42; DB 6; Length 475;
Best Local Similarity 72.7%; Pred. No. 1.8e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPGE 11
||: ||| |||
Db 49 LPSYVGLHPGE 59

RESULT 10
AAE34729
ID AAE34729 standard; protein; 475 AA.
XX
AC AAE34729;
XX
DT 14-MAY-2003 (first entry)
XX
DE Streptomyces albofaciens ema8 protein.
XX
KW P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;
KW emamectin; insecticide; ema8 protein.
XX
OS Streptomyces albofaciens.
XX
PN WO200292801-A2.
XX
PD 21-NOV-2002.
XX
PF 15-MAY-2002; 2002WO-EP005363.
XX
PR 16-MAY-2001; 2001US-0291149P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
PI Buckel TG;
XX
DR WPI; 2003-140280/13.
DR N-PSDB; AAD53024.
XX
PT Novel polypeptide that exhibits enzymatic activity of P450 monooxygenase
PT and capable of regioselectively oxidizing alcohol of avermectin useful
PT for making emamectin from avermectin.
XX
PS Claim 17; Page 116-117; 157pp; English.
XX
CC The present invention relates to novel proteins that exhibit an enzymatic
CC activity of P450 monooxygenase and capable of regioselectively oxidising
CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
CC The invention also relates to ferrodoxins and ferrodoxin reductases that
CC are active with the P450 monooxygenases. Sequences of the invention are
CC useful for producing 4-keto-avermectin from avermectin, which is useful
CC for producing emamectin. Emamectin is useful as an insecticide. The
CC present sequence is Streptomyces albofaciens ema8 protein
XX
SQ Sequence 475 AA;

Query Match 64.6%; Score 42; DB 6; Length 475;
Best Local Similarity 72.7%; Pred. No. 1.8e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPGE 11
||: ||| |||
Db 49 LPSYVGLHPGE 59

RESULT 11
ABJ10397
ID ABJ10397 standard; protein; 106 AA.
XX
AC ABJ10397;
XX
DT 28-NOV-2002 (first entry)
XX
DE Mutant anti-mesothelin Fv (ST6) variable light chain.
XX
KW Mutant; mutein; recombinant immunotoxin; reduced liver toxicity; cancer.
XX
OS Unidentified.
OS Synthetic.
XX
PN WO200240545-A2.
XX
PD 23-MAY-2002.
XX
PF 16-NOV-2001; 2001WO-US043602.
XX
PR 17-NOV-2000; 2000US-0249805P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Pastan IH, Onda M, Nagata S, Tsutsumi Y, Vincent JJ, Kreitman RJ;
PI Vasmatzis G, Lee B;
XX
DR WPI; 2002-500208/53.
DR N-PSDB; ABT08085.
XX
PT Recombinant immunotoxins with reduced liver toxicity for killing
PT malignant cells bearing antigen comprising mutated framework regions of
PT antibody heavy and/or light chain or antigen-binding portion of
PT immunotoxin.
XX
PS Claim 12; Fig 2; 82pp; English.
XX

```
CC   The invention comprises the amino acid and coding sequences of
CC   recombinant immunotoxin proteins. The immunotoxin proteins of the
CC   invention contain an antibody (or an antigen-binding fragment) with a
CC   substitution of a negatively charged amino acid for an uncharged or
CC   positively charged amino acid. The immunotoxins of the invention have
CC   reduced liver toxicity. The immunotoxins of the invention are useful for
CC   killing a malignant cell (e.g. a cancer cell). The present amino acid
CC   sequence represents a recombinant immunotoxin of the invention
XX
SQ   Sequence 106 AA;

  Query Match             63.1%;  Score 41;  DB 5;  Length 106;
  Best Local Similarity   63.6%;  Pred. No. 51;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 PAVVGLSPGEQ 12
              ||::  |||||
Db          8 PAIMSASPGEQ 18


RESULT 12
ABJ10395
ID   ABJ10395 standard; protein; 106 AA.
XX
AC   ABJ10395;
XX
DT   28-NOV-2002  (first entry)
XX
DE   Mutant anti-Tac Fv (M16) variable light chain.
XX
KW   Mutant; mutein; recombinant immunotoxin; reduced liver toxicity; cancer.
XX
OS   Unidentified.
OS   Synthetic.
XX
PN   WO200240545-A2.
XX
PD   23-MAY-2002.
XX
PF   16-NOV-2001; 2001WO-US043602.
XX
PR   17-NOV-2000; 2000US-0249805P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Pastan IH,  Onda M,  Nagata S,  Tsutsumi Y,  Vincent JJ,  Kreitman RJ;
PI   Vasmatzis G,  Lee B;
XX
DR   WPI; 2002-500208/53.
DR   N-PSDB; ABT08083.
XX
PT   Recombinant immunotoxins with reduced liver toxicity for killing
PT   malignant cells bearing antigen comprising mutated framework regions of
PT   antibody heavy and/or light chain or antigen-binding portion of
PT   immunotoxin.
XX
PS   Claim 8; Fig 2; 82pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of
CC   recombinant immunotoxin proteins. The immunotoxin proteins of the
CC   invention contain an antibody (or an antigen-binding fragment) with a
CC   substitution of a negatively charged amino acid for an uncharged or
CC   positively charged amino acid. The immunotoxins of the invention have
CC   reduced liver toxicity. The immunotoxins of the invention are useful for
CC   killing a malignant cell (e.g. a cancer cell). The present amino acid
CC   sequence represents a recombinant immunotoxin of the invention
XX
SQ   Sequence 106 AA;

  Query Match             63.1%;  Score 41;  DB 5;  Length 106;
  Best Local Similarity   63.6%;  Pred. No. 51;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 PAVVGLSPGEQ 12
              ||::  |||||
Db          8 PAIMSASPGEQ 18


RESULT 13
ADD49105
ID   ADD49105 standard; protein; 530 AA.
XX
AC   ADD49105;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Human NOV15k SEQ ID 78.
XX
KW   Antidiabetic; anorectic; cardiant; hypotensive; antiarteriosclerotic;
KW   virucide; antibacterial; fungicide; protozoacide; nootropic;
KW   neuroprotective; antiparkinsonian; anticonvulsant; osteopathic;
KW   antiarthritic; antiinflammatory; dermatological; antiasthmatic;
KW   antilipemic; gene therapy; NOV protein; metabolic disorder; diabetes;
KW   obesity; viral infection; bacterial infection; fungal infection;
KW   helminthic infection; protozoal infection; anorexia; cancer;
KW   cardiovascular disease; hypertension; atherosclerosis;
KW   neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW   epilepsy; immune disorder; osteoarthritis; haematopoietic disorder;
KW   inflammatory skin disorder; asthma; dyslipidemia; human.
XX
OS   Homo sapiens.
XX
PN   WO2003060149-A2.
XX
PD   24-JUL-2003.
XX
PF   06-JAN-2003; 2003WO-US000252.
XX
PR   04-JAN-2002; 2002US-0345222P.
PR   14-JAN-2002; 2002US-0348693P.
PR   16-JAN-2002; 2002US-0349182P.
PR   17-JAN-2002; 2002US-0349733P.
PR   18-JAN-2002; 2002US-0350263P.
PR   24-JAN-2002; 2002US-0351977P.
PR   28-MAY-2002; 2002US-0383758P.
PR   05-JUN-2002; 2002US-0385969P.
PR   11-JUN-2002; 2002US-0387834P.
PR   17-JUL-2002; 2002US-0396407P.
PR   30-SEP-2002; 2002US-0415115P.
PR   03-JAN-2003; 2003US-00336603.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Grosse WM,  Alsobrook JP,  Anderson DW,  Burgess CE,  Edinger SR;
PI   Ellerman K,  Furtak K,  Gangolli EA,  Gerlach VL,  Gilbert JA;
PI   Gunther E,  Gorman L,  Guo X,  Ji W,  Li L,  Miller CE,  Padigaru M;
PI   Patturajan M,  Rastelli L,  Macdougall JR,  Mishra VS,  Smithson G;
PI   Spytek KA,  Stone DJ,  Shenoy SG,  Taupier RJ,  Vernet CAM,  Zhong M;
PI   Malyankar UM,  Millet I,  Kekuda R;
XX
DR   WPI; 2003-587288/55.
DR   N-PSDB; ADD49104.
XX
PT   New isolated NOVX polypeptides and polynucleotides, useful for
PT   preventing, diagnosing or treating NOVX-associated disorders, e.g.
PT   osteoarthritis, obesity, atherosclerosis, cancer, Parkinson's disease,
PT   asthma, or infections.
XX
PS   Claim 1; Page 181-182; 311pp; English.
XX
CC   The present invention relates to novel NOV proteins and their coding
CC   sequences (ADD49028-ADD49131). The proteins and coding sequences are
CC   useful in the manufacture of a medicament for treating a syndrome
CC   associated with a human disease, preferably a NOV-associated disorder
CC   such as metabolic disorders, diabetes, obesity, infectious diseases
CC   (viral, bacterial, fungal, helminthic, and protozoal), anorexia, cancer,
```

```
CC   cardiovascular diseases (hypertension, atherosclerosis),
CC   neurodegenerative disorders, Alzheimer's disease, Parkinson's disease,
CC   epilepsy, immune disorders (osteoarthritis), hematopoietic disorders,
CC   inflammatory skin disorders, asthma and various dyslipidemias. The coding
CC   sequences and proteins may also be used as targets for the identification
CC   of small molecules that modulate or inhibit e.g. neurogenesis, cell
CC   differentiation, cell proliferation, haematopoiesis, wound healing and
CC   angiogenesis, in gene therapy, in generation of antibodies that bind
CC   immunospecifically to NOV substances for use in therapeutic or diagnostic
CC   methods.
XX
SQ   Sequence 530 AA;

  Query Match             63.1%;  Score 41;  DB 7;  Length 530;
  Best Local Similarity   75.0%;  Pred. No. 2.9e+02;
  Matches    9;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQ 12
              || ||| || ||
Db        416 LPPVVGGSPAEQ 427


RESULT 14
ADD49091
ID   ADD49091 standard; protein; 549 AA.
XX
AC   ADD49091;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Human NOV15d SEQ ID 64.
XX
KW   Antidiabetic; anorectic; cardiant; hypotensive; antiarteriosclerotic;
KW   virucide; antibacterial; fungicide; protozoacide; nootropic;
KW   neuroprotective; antiparkinsonian; anticonvulsant; osteopathic;
KW   antiarthritic; antiinflammatory; dermatological; antiasthmatic;
KW   antilipemic; gene therapy; NOV protein; metabolic disorder; diabetes;
KW   obesity; viral infection; bacterial infection; fungal infection;
KW   helminthic infection; protozoal infection; anorexia; cancer;
KW   cardiovascular disease; hypertension; atherosclerosis;
KW   neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW   epilepsy; immune disorder; osteoarthritis; haematopoietic disorder;
KW   inflammatory skin disorder; asthma; dyslipidemia; human.
XX
OS   Homo sapiens.
XX
PN   WO2003060149-A2.
XX
PD   24-JUL-2003.
XX
PF   06-JAN-2003; 2003WO-US000252.
XX
PR   04-JAN-2002; 2002US-0345222P.
PR   14-JAN-2002; 2002US-0348693P.
PR   16-JAN-2002; 2002US-0349182P.
PR   17-JAN-2002; 2002US-0349733P.
PR   18-JAN-2002; 2002US-0350263P.
PR   24-JAN-2002; 2002US-0351977P.
PR   28-MAY-2002; 2002US-0383758P.
PR   05-JUN-2002; 2002US-0385969P.
PR   11-JUN-2002; 2002US-0387834P.
PR   17-JUL-2002; 2002US-0396407P.
PR   30-SEP-2002; 2002US-0415115P.
PR   03-JAN-2003; 2003US-00336603.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Grosse WM,  Alsobrook JP,  Anderson DW,  Burgess CE,  Edinger SR;
PI   Ellerman K,  Furtak K,  Gangolli EA,  Gerlach VL,  Gilbert JA;
PI   Gunther E,  Gorman L,  Guo X,  Ji W,  Li L,  Miller CE,  Padigaru M;
PI   Patturajan M,  Rastelli L,  Macdougall JR,  Mishra VS,  Smithson G;
PI   Spytek KA,  Stone DJ,  Shenoy SG,  Taupier RJ,  Vernet CAM,  Zhong M;
PI   Malyankar UM,  Millet I,  Kekuda R;
XX
DR   WPI; 2003-587288/55.
DR   N-PSDB; ADD49090.
XX
PT   New isolated NOVX polypeptides and polynucleotides, useful for
PT   preventing, diagnosing or treating NOVX-associated disorders, e.g.
PT   osteoarthritis, obesity, atherosclerosis, cancer, Parkinson's disease,
PT   asthma, or infections.
XX
PS   Claim 1; Page 177; 311pp; English.
XX
CC   The present invention relates to novel NOV proteins and their coding
CC   sequences (ADD49028-ADD49131). The proteins and coding sequences are
CC   useful in the manufacture of a medicament for treating a syndrome
CC   associated with a human disease, preferably a NOV-associated disorder
CC   such as metabolic disorders, diabetes, obesity, infectious diseases
CC   (viral, bacterial, fungal, helminthic, and protozoal), anorexia, cancer,
CC   cardiovascular diseases (hypertension, atherosclerosis),
CC   neurodegenerative disorders, Alzheimer's disease, Parkinson's disease,
CC   epilepsy, immune disorders (osteoarthritis), hematopoietic disorders,
CC   inflammatory skin disorders, asthma and various dyslipidemias. The coding
CC   sequences and proteins may also be used as targets for the identification
CC   of small molecules that modulate or inhibit e.g. neurogenesis, cell
CC   differentiation, cell proliferation, haematopoiesis, wound healing and
CC   angiogenesis, in gene therapy, in generation of antibodies that bind
CC   immunospecifically to NOV substances for use in therapeutic or diagnostic
CC   methods.
XX
SQ   Sequence 549 AA;

  Query Match             63.1%;  Score 41;  DB 7;  Length 549;
  Best Local Similarity   75.0%;  Pred. No. 3e+02;
  Matches    9;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 LPAVVGLSPGEQ 12
              || ||| || ||
Db        419 LPPVVGGSPAEQ 430


RESULT 15
ABG70277
ID   ABG70277 standard; protein; 579 AA.
XX
AC   ABG70277;
XX
DT   05-NOV-2002  (first entry)
XX
DE   Human Glypican-2 Precursor-like protein #1.
XX
KW   Human; NOVX; pathological condition; NOVX-associated disorder;
KW   Von Hippel-Lindau syndrome; cirrhosis; transplantation disorder;
KW   pancreatitis; obesity; diabetes; autoimmune disease; infertility;
KW   renal artery stenosis; interstitial nephritis; glomerulonephritis;
KW   polycystic kidney disease; cataract; Alzheimer's disease; cancer;
KW   acoustic trauma; cardiomyopathy; atherosclerosis; hypertension;
KW   congenital heart defect; scleroderma; endometriosis; haemophilia;
KW   dementia; stroke; Parkinson's disease; Huntington's disease; epilepsy;
KW   multiple sclerosis; anxiety; pain; leukaemia; hypothyroidism; psoriasis;
KW   acne; wound; asthma; human disease; calpain; epsin; zinc finger;
KW   low density lipoprotein B; LDLB; purinoceptor; CG8841; synaptotagmin;
KW   serine protease TLSP; mitogen activated protein kinase kinase-2;
KW   glypican-2 precursor; thymosin beta-10.
XX
OS   Homo sapiens.
XX
PN   WO200255702-A2.
XX
PD   18-JUL-2002.
XX
PF   26-OCT-2001; 2001WO-US050925.
XX
PR   26-OCT-2000; 2000US-0243320P.
PR   26-OCT-2000; 2000US-0243592P.
```

PR 26-OCT-2000; 2000US-0243642P.
PR 27-OCT-2000; 2000US-0243681P.
PR 27-OCT-2000; 2000US-0243863P.
PR 31-OCT-2000; 2000US-0244443P.
PR 01-NOV-2000; 2000US-0244995P.
PR 01-NOV-2000; 2000US-0245029P.
PR 02-NOV-2000; 2000US-0245293P.
PR 02-NOV-2000; 2000US-0245315P.
PR 02-NOV-2000; 2000US-0245316P.
PR 19-JAN-2001; 2001US-0262994P.
PR 15-FEB-2001; 2001US-0269056P.
PR 02-MAR-2001; 2001US-0272923P.
PR 15-MAR-2001; 2001US-0276565P.
PR 07-SEP-2001; 2001US-0318119P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Gangolli EA, Spytek KA, Gilbert J, Casman S, Blalock A, Li L;
PI Vernet CAM, Shenoy S, Mishra V, Furtak K, Gerlach V, Edinger S;
PI Malyankar U, Stone D, Millet I, Smithson G, Gunther E, Padigaru M;
PI Taupier RJ, Anderson D;
XX
DR WPI; 2002-590673/63.
DR N-PSDB; ABK51684.
XX
PT Isolated NOVX polypeptides and nucleic acid molecules useful for
PT treating, preventing, diagnosing and researching pathological conditions
PT in humans with a NOVX-associated disorders, e.g. cancer, stroke or
PT Alzheimer's disease.
XX
PS Claim 1; Page 57; 236pp; English.
XX
CC The present invention relates to a new polypeptide that comprises any of
CC 17 fully defined sequences of 43-990 amino acids given in the
CC specification. The NOVX polypeptide, nucleic acid and antibody of the
CC invention are useful for treating or preventing a pathological condition
CC in humans with a NOVX-associated disorder, e.g. Von Hippel-Lindau
CC syndrome, cirrhosis, transplantation disorders, pancreatitis, obesity,
CC diabetes, autoimmune disease, renal artery stenosis, interstitial
CC nephritis, glomerulonephritis, polycystic kidney disease, cataract,
CC Alzheimer's disease, acoustic trauma, cancer, infertility,
CC cardiomyopathies, atherosclerosis, hypertension, congenital heart
CC defects, scleroderma, endometriosis, haemophilia, dementia, stroke,
CC Parkinson's disease, Huntington's disease, epilepsy, multiple sclerosis,
CC anxiety, pain, leukaemias, hypothyroidism, psoriasis, acne, wounds and
CC asthma. They are also useful for the manufacture of a medicament for
CC treating a syndrome associated with a human disease, specifically a NOVX-
CC associated disorder. They may also be useful in therapeutic applications
CC including protein therapy, as small molecule drug targets, as antibody
CC targets, as diagnostic and/or prognostic markers, in gene therapy, as
CC research tools and in tissue regeneration. The present amino acid
CC sequence represents one of the 17 novel proteins of the invention
XX
SQ Sequence 579 AA;

Query Match 63.1%; Score 41; DB 5; Length 579;
Best Local Similarity 75.0%; Pred. No. 3.2e+02;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 LPAVVGLSPGEQ 12
              || ||| || ||
Db        439 LPPVVGGSPAEQ 450
```

Search completed: May 7, 2004, 12:33:46
Job time : 36.27 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:20:30 ; Search time 5.4 Seconds
(without alignments)
144.639 Million cell updates/sec

Title: US-09-786-214A-15
Perfect score: 75
Sequence: 1 AGLPAVVGLSPGEQE 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	57.3	658	1	VG18_BPT4	P13332 bacteriopha
2	42	56.0	326	1	MER_METTI	Q9uxp0 methanolobu
3	42	56.0	508	1	COBI_MYCTU	Q10677 mycobacteri
4	41	54.7	156	1	RUVX_CAUCR	Q9a5k8 caulobacter
5	41	54.7	213	1	MDCG_XANAC	Q8ppw9 xanthomonas
6	41	54.7	520	1	AT15_YEAST	P25641 saccharomyc
7	41	54.7	774	1	LOL2_HUMAN	Q9y4k0 homo sapien
8	40	53.3	157	1	RISB_PYRFU	Q8u418 pyrococcus
9	40	53.3	359	1	ALF2_PEA	P46257 pisum sativ
10	40	53.3	359	1	ALF_CICAR	O65735 cicer ariet
11	40	53.3	361	1	VAL1_TMOV	Q06657 tomato mott
12	40	53.3	633	1	PLB5_SCHPO	Q9y7n6 schizosacch
13	40	53.3	773	1	YHGF_ECOLI	P46837 escherichia
14	40	53.3	813	1	CADM_MOUSE	Q9wtp5 mus musculu
15	40	53.3	2364	1	PGCA_BOVIN	P13608 bos taurus
16	39	52.0	115	1	KV3I_HUMAN	P04433 homo sapien
17	39	52.0	115	1	KV5I_MOUSE	P01642 mus musculu
18	39	52.0	146	1	DTD_XANAC	Q8pg32 xanthomonas
19	39	52.0	385	1	YG53_METJA	Q59047 methanococc
20	39	52.0	597	1	NR41_RAT	P22829 rattus norv
21	39	52.0	1402	1	N160_MOUSE	Q9z0w3 mus musculu
22	39	52.0	1596	1	GLI3_HUMAN	P10071 homo sapien
23	39	52.0	1636	1	BUD3_YEAST	P25558 saccharomyc
24	38	50.7	146	1	DTD_XANCP	Q8p4h1 xanthomonas
25	38	50.7	299	1	DAP1_STRCO	O86841 streptomyce
26	38	50.7	299	1	NUCG_BOVIN	P38447 bos taurus
27	38	50.7	331	1	ILVC_ANASP	Q8yum5 anabaena sp
28	38	50.7	344	1	ILVC_OCEIH	Q8en66 oceanobacil
29	38	50.7	390	1	COBL_MYCTU	Q10671 mycobacteri
30	38	50.7	395	1	HMEB_ARCFU	O29750 archaeoglob
31	38	50.7	429	1	RS1_LEULA	P50889 leuconostoc
32	38	50.7	484	1	COBQ_PSEDE	P29932 pseudomonas
33	38	50.7	507	1	CATA_PICAN	P30263 pichia angu
34	38	50.7	516	1	C4AD_DROME	Q9v4t3 drosophila
35	38	50.7	628	1	HNFA_MOUSE	P22361 mus musculu
36	38	50.7	628	1	HNFA_RAT	P15257 rattus norv
37	38	50.7	631	1	HNFA_HUMAN	P20823 homo sapien
38	38	50.7	699	1	EFG_HAEIN	P43925 haemophilus
39	38	50.7	700	1	EFG_PASMU	P57938 pasteurella
40	38	50.7	876	1	TOPI_VIBCH	Q9krb2 vibrio chol
41	38	50.7	895	1	ANDR_MACFA	O97952 macaca fasc
42	38	50.7	895	1	ANDR_PAPHA	O97960 papio hamad
43	38	50.7	899	1	ANDR_MOUSE	P19091 mus musculu
44	38	50.7	902	1	ANDR_RAT	P15207 rattus norv
45	38	50.7	907	1	ANDR_CANFA	Q9tt90 canis famil

ALIGNMENTS

```
RESULT 1
VG18_BPT4
ID   VG18_BPT4      STANDARD;      PRT;   658 AA.
AC   P13332; Q9T0U3;
DT   01-JAN-1990 (Rel. 13, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Tail sheath protein Gp18.
GN   18.
OS   Bacteriophage T4.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC   T4-like viruses.
OX   NCBI_TaxID=10665;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=88155753; PubMed=2964531;
RA   Arisaka F., Nakako T., Takahashi H., Ishii S.-I.;
RT   "Nucleotide sequence of the tail sheath gene of bacteriophage T4 and
RT   amino acid sequence of its product.";
RL   J. Virol. 62:1186-1193(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22514363; PubMed=12626685;
RA   Miller E.S., Kutter E., Mosig G., Arisaka F., Kunisawa T., Ruger W.;
RT   "Bacteriophage T4 genome.";
RL   Microbiol. Mol. Biol. Rev. 67:86-156(2003).
RN   [3]
RP   SEQUENCE OF 638-658 FROM N.A.
RX   MEDLINE=88119228; PubMed=2963141;
RA   Arisaka F., Ishimoto L., Kassavetis G., Kumazaki T., Ishii S.-I.;
RT   "Nucleotide sequence of the tail tube structural gene of
RT   bacteriophage T4.";
RL   J. Virol. 62:882-886(1988).
CC   -!- FUNCTION: The contractile tail of bacteriophage T4 consists of a
CC       contractile sheath, a tube and a baseplate. 144 protomers of Gp18,
CC       arranged in 24 annuli, form the contractile tail sheath.
CC   -!- SUBUNIT: Monomer.
CC   -!- SIMILARITY: 30% IDENTITY TO THE TAIL SHEATH PROTEIN OF PHAGE PS17,
CC       AND BY EXTENSION, TO P.AERUGINOSA R-TYPE PYOCINS (ANTIBIOTICS).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M19085; AAA32541.1; -.
DR   EMBL; AF158101; AAD42423.1; -.
DR   PIR; JF0021; GKBPT4.
DR   InterPro; IPR007067; Phage_sheath_1.
DR   Pfam; PF04984; Phage_sheath_1; 1.
KW   Structural protein.
FT   INIT_MET      0      0
FT   DOMAIN      171    175       SER-RICH.
```

```
FT   CONFLICT     99     99       D -> E (IN REF. 1).
FT   CONFLICT    147    150       GKNY -> AKII (IN REF. 1).
FT   CONFLICT    300    300       E -> G (IN REF. 1).
FT   CONFLICT    398    398       A -> V (IN REF. 1).
FT   CONFLICT    453    453       H -> Y (IN REF. 1).
FT   CONFLICT    594    594       N -> I (IN REF. 1).
SQ   SEQUENCE   658 AA;  71199 MW;  6E045F40D39AF21D CRC64;

  Query Match             57.3%;  Score 43;  DB 1;  Length 658;
  Best Local Similarity   66.7%;  Pred. No. 34;
  Matches    8;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 GLPAVVGLSPGE 13
              |:| || | |||
Db        217 GIPGVVALYPGE 228


RESULT 2
MER_METTI
ID   MER_METTI      STANDARD;      PRT;   326 AA.
AC   Q9UXP0;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Coenzyme F420-dependent N(5),N(10)-methylenetetrahydromethanopterin
DE   reductase (EC 1.5.99.11) (Methylene-H(4)MPT reductase).
GN   MER OR FFDA.
OS   Methanolobus tindarius.
OC   Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC   Methanosarcinaceae; Methanolobus.
OX   NCBI_TaxID=2221;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 2278;
RX   MEDLINE=99132696; PubMed=9933933;
RA   Westenberg D.J., Braune A., Ruppert C., Mueller V., Herzberg C.,
RA   Gottschalk G., Blaut M.;
RT   "The F420H2-dehydrogenase from Methanolobus tindarius: cloning of the
RT   ffd operon and expression of the genes in Escherichia coli.";
RL   FEMS Microbiol. Lett. 170:389-398(1999).
CC   -!- FUNCTION: Catalyzes the reversible reduction of methylene-H(4)MPT
CC       to methyl-H(4)MPT (By similarity).
CC   -!- CATALYTIC ACTIVITY: N(5),N(10)-methylenetetrahydromethanopterin +
CC       reduced coenzyme F420 = 5-methyl-5,6,7,8-tetrahydromethanopterin +
CC       coenzyme F420.
CC   -!- PATHWAY: Methanogenesis from carbon dioxide; fifth step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the mer family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AJ011519; CAB56639.1; -.
DR   PIR; T45226; T45226.
DR   HAMAP; MF_01091; -; 1.
DR   InterPro; IPR002103; Bac_luciferase.
DR   Pfam; PF00296; bac_luciferase; 1.
KW   Methanogenesis; One-carbon metabolism; Oxidoreductase.
SQ   SEQUENCE   326 AA;  34043 MW;  16F3AB9733A45D82 CRC64;

  Query Match             56.0%;  Score 42;  DB 1;  Length 326;
  Best Local Similarity   57.1%;  Pred. No. 24;
  Matches    8;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 AGLPAVVGLSPGEQ 14
              :|  |::|| ||||
Db         84 SGGRAILGLGPGEQ 97


RESULT 3
COBI_MYCTU
ID   COBI_MYCTU     STANDARD;      PRT;   508 AA.
AC   Q10677;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Cobalamin biosynthesis protein COBIJ [Includes: Precorrin-2 C20-
DE   methyltransferase (EC 2.1.1.130) (S-adenosyl-L-methionine--precorrin-2
DE   methyltransferase) (SP2MT); Precorrin-3 methylase (EC 2.1.1.-)].
GN   COBIJ OR COBI OR RV2066 OR MT2126 OR MTCY49.05 OR MB2092.
OS   Mycobacterium tuberculosis, and
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773, 1765;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX   MEDLINE=22206494; PubMed=12218036;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT   "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   J. Bacteriol. 184:5479-5490(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.bovis; STRAIN=AF2122/97;
RX   MEDLINE=22709107; PubMed=12788972;
RA   Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA   Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA   Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA   Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT   "The complete genome sequence of Mycobacterium bovis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC   -!- FUNCTION: METHYLATES PRECORRIN-2 AT THE C-20 POSITION TO PRODUCE
CC       PRECORRIN-3A (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-2 = S-
CC       adenosyl-L-homocysteine + precorrin-3A.
CC   -!- PATHWAY: Cobalamin biosynthesis.
CC   -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS SUMT, CYSG, CBIF/COBM
CC       AND CBIL/COBI.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z73966; CAA98214.1; -.
DR   EMBL; AE007063; AAK46406.1; -.
DR   EMBL; BX248341; CAD96945.1; -.
```

```
DR   PIR; E70764; E70764.
DR   TIGR; MT2126; -.
DR   TubercuList; Rv2066; -.
DR   InterPro; IPR006364; CobI_CbiL.
DR   InterPro; IPR006363; CobJ.
DR   InterPro; IPR000878; Cor/por_Metransf.
DR   InterPro; IPR003043; Uropor_Metransf.
DR   Pfam; PF00590; TP_methylase; 2.
DR   TIGRFAMs; TIGR01467; cobI_cbiL; 1.
DR   TIGRFAMs; TIGR01466; cobJ_cbiH; 1.
DR   PROSITE; PS00839; SUMT_1; 1.
DR   PROSITE; PS00840; SUMT_2; 1.
KW   Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW   Methyltransferase; Multifunctional enzyme; Complete proteome.
FT   DOMAIN       1    243       PRECORRIN-2 C20-METHYLTRANSFERASE.
FT   DOMAIN     244    508       PRECORRIN-3 METHYLASE.
SQ   SEQUENCE   508 AA;  53910 MW;  95AC066F022C4DC1 CRC64;

  Query Match             56.0%;  Score 42;  DB 1;  Length 508;
  Best Local Similarity   57.1%;  Pred. No. 38;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 GLPAVVGLSPGEQE 15
              |  ||||| ||: :
Db        247 GTVAVVGLGPGDSD 260


RESULT 4
RUVX_CAUCR
ID   RUVX_CAUCR     STANDARD;      PRT;   156 AA.
AC   Q9A5K8;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Putative Holliday junction resolvase (EC 3.1.-.-).
GN   CC2439.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC   Caulobacteraceae; Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC   -!- FUNCTION: Could be a nuclease that resolves Holliday junction
CC       intermediates in genetic recombination.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: Belongs to the yqgF HJR family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE005913; AAK24410.1; -.
DR   PIR; F87551; F87551.
DR   TIGR; CC2439; -.
DR   HAMAP; MF_00651; -; 1.
DR   InterPro; IPR005227; Cons_hypoth250.
DR   InterPro; IPR006641; YqgFc.
DR   Pfam; PF03652; UPF0081; 1.
DR   SMART; SM00732; YqgFc; 1.
DR   TIGRFAMs; TIGR00250; TIGR00250; 1.
KW   Hydrolase; Nuclease; DNA repair; DNA recombination; Complete proteome.
SQ   SEQUENCE   156 AA;  17142 MW;  21F54D8648396141 CRC64;

  Query Match             54.7%;  Score 41;  DB 1;  Length 156;
  Best Local Similarity   80.0%;  Pred. No. 17;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          5 AVVGLSPGEQ 14
              ||||| |||:
Db         18 AVVGLDPGEK 27


RESULT 5
MDCG_XANAC
ID   MDCG_XANAC     STANDARD;      PRT;   213 AA.
AC   Q8PPW9;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phosphoribosyl-dephospho-CoA transferase (EC 2.7.7.-) (Holo-ACP
DE   synthase).
GN   MDCG OR XAC0564.
OS   Xanthomonas axonopodis (pv. citri).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xanthomonas.
OX   NCBI_TaxID=92829;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=306 / ATCC 13902 / XV 101;
RX   MEDLINE=22022145; PubMed=12024217;
RA   da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA   Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA   Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA   Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA   Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA   Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA   Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA   Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA   Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA   Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA   Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA   Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA   Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA   Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA   Setubal J.C., Kitajima J.P.;
RT   "Comparison of the genomes of two Xanthomonas pathogens with differing
RT   host specificities.";
RL   Nature 417:459-463(2002).
CC   -!- FUNCTION: Transfers 2-(5''-triphosphoribosyl)-3'-
CC       dephosphocoenzyme-A to the apo-acyl carrier protein of the
CC       malonate decarboxylase to yield holo-acyl carrier protein (By
CC       similarity).
CC   -!- CATALYTIC ACTIVITY: 2'-(5''-triphosphoribosyl)-3'-dephospho-CoA +
CC       apo-ACP = holo-ACP + diphosphate.
CC   -!- SIMILARITY: Belongs to the mdcG family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE011683; AAM35453.1; -.
DR   HAMAP; MF_00650; -; 1.
KW   Transferase; Nucleotidyltransferase; Complete proteome.
FT   ACT_SITE    135    135       BY SIMILARITY.
FT   ACT_SITE    137    137       BY SIMILARITY.
SQ   SEQUENCE   213 AA;  23013 MW;  C378D8A975C75A17 CRC64;
```

```
  Query Match             54.7%;  Score 41;  DB 1;  Length 213;
  Best Local Similarity   64.3%;  Pred. No. 23;
  Matches    9;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 AGLPAVVGLSPGEQ 14
              ||||||     | |
Db         33 AGLPAVVARGDGSQ 46


RESULT 6
AT15_YEAST
ID   AT15_YEAST     STANDARD;      PRT;   520 AA.
AC   P25641; Q8NIL6;
DT   01-MAY-1992 (Rel. 22, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Autophagy related putative lipase ATG15 (EC 3.1.1.3) (Cytoplasm to
DE   vacuole targeting protein 17).
GN   ATG15 OR AUT5 OR CVT17 OR YCR068W OR YCR68W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, GLYCOSYLATION, AND MUTAGENESIS OF
RP   SER-332.
RX   MEDLINE=21125771; PubMed=11085977;
RA   Teter S.A., Eggerton K.P., Scott S.V., Kim J., Fischer A.M.,
RA   Klionsky D.J.;
RT   "Degradation of lipid vesicles in the yeast vacuole requires function
RT   of Cvt17, a putative lipase.";
RL   J. Biol. Chem. 276:2083-2087(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=92244356; PubMed=1574125;
RA   Oliver S.G., van der Aart Q.J.M., Agostoni-Carbone M.L., Aigle M.,
RA   Alberghina L., Alexandraki D., Antoine G., Anwar R., Ballesta J.P.G.,
RA   Benit P., Berben G., Bergantino E., Biteau N., Bolle P.-A.,
RA   Bolotin-Fukuhara M., Brown A.J.P., Brown R., Buhler J.-M.,
RA   Carignani G., Chanet R., Contreras R., Crouzet M., Daignan-Fornier B.,
RA   De Haan M., Defoor E., Delgado M.D., Demolder J., Doira C., Dubois E.,
RA   Dujon B., Duesterhoeft A., Erdmann D., Esteban M., Fabre F.,
RA   Fairhead C.A., Faye G., Feldmann H., Fiers W.,
RA   Francingues-Gaillard M.-C., Franco L., Frontali L., Fukuhara H.,
RA   Fuller L.J., Gent M.E., Gigot D., Gilliquet V., Glansdorff N.,
RA   Goffeau A., Grenson M., Grisanti P., Grivell L.A., Haasemann M.,
RA   Hatat D., Hegemann J.H., Herbert C.J., Hilger F., Hohmann S.,
RA   Hollenberg C.P., Huse K., Iborra F., Indge K.J., Isono K., Jackman P.,
RA   Jacq C., Jacquet M., James C.M., Jauniaux J.-C., Jia Y., Jimenez A.,
RA   Kleinhans U., Kreisl P., Lafranchi G., Lewis C., van der Linden C.G.,
RA   Lucchini G., Lutzenkirchen K., Maat C., Mannhaupt G., Manzano M.E.,
RA   Martegani E., Mathieu A., Maurer C.T.C., McConnell D., McKee R.A.,
RA   Messenguy F., Mewes H.-W., Molemans F., Montague M.A., Navas L.,
RA   Newlon C.S., Olson M.V., Pallier C., Panzeri L., Pearson B.M.,
RA   Perea J., Philippsen P., Pierard A., Planta R.J., Plevani P.,
RA   Poetsch B., Pohl F.M., Purnelle B., Ramezani Rad M., Rasmussen S.W.,
RA   Raynal A., Remacha M., Richterich P., Roberts A.B., Rodriguez F.,
RA   Sanz E., Schaaff-Gerstenschlaeger I., Scherens B., Schweitzer B.,
RA   Shu Y., Skala J., Slonimski P.P., Sor F., Soustelle C.,
RA   Spiegelberg R., Stateva L.I., Steensma H.Y., Steiner S., Thierry A.,
RA   Thireos G., Triano L.N., Urrestarazu L.A., Valle G., Vetter I.,
RA   van Vliet-Reedijk J.C., Volckaert G., Vreken P., Warmington J.R.,
RA   von Wettstein D., Wicksteed B.L., Wilson C., Wurst H., Xu G.,
RA   Zimmermann F.K., Sgouros J.G.;
RT   "The complete DNA sequence of yeast chromosome III.";
RL   Nature 357:38-46(1992).
RN   [3]
RP   REVISIONS TO C-TERMINUS.
RA   Valles G., Volckaerts G.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   FUNCTION, SUBCELLULAR LOCATION, GLYCOSYLATION, AND MUTAGENESIS OF
RP   SER-332.
RX   MEDLINE=21450820; PubMed=11566994;
RA   Epple U.D., Suriapranata I., Eskelinen E.-L., Thumm M.;
RT   "Aut5/Cvt17p, a putative lipase essential for disintegration of
RT   autophagic bodies inside the vacuole.";
RL   J. Bacteriol. 183:5942-5955(2001).
RN   [5]
RP   FUNCTION, SUBCELLULAR LOCATION, TOPOLOGY, AND GLYCOSYLATION.
RX   MEDLINE=22499636; PubMed=12499386;
RA   Epple U.D., Eskelinen E.-L., Thumm M.;
RT   "Intravacuolar membrane lysis in Saccharomyces cerevisiae. Does
RT   vacuolar targeting of Cvt17/Aut5p affect its function?";
RL   J. Biol. Chem. 278:7810-7821(2003).
CC   -!- FUNCTION: Essential for lysis of subvacuolar cytoplasm to vacuole
CC       targeted bodies and intravacuolar autophagic bodies. Involved in
CC       the lysis of intravacuolar multivesicular body (MVB) vesicles.
CC   -!- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a
CC       fatty acid anion.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
CC       reticulum. From ER, targeted to vacuolar lumen at the MVB vesicles
CC       via the Golgi and the prevacuolar compartment (PVC).
CC   -!- PTM: Glycosylated.
CC   -!- SIMILARITY: Belongs to the AB hydrolase superfamily. Lipase
CC       family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X59720; CAC42987.1; -.
DR   PIR; S19483; S19483.
DR   PIR; S74292; S74292.
DR   GermOnline; 138970; -.
DR   SGD; S0000664; ATG15.
DR   GO; GO:0016021; C:integral to membrane; IPI.
DR   GO; GO:0005775; C:vacuolar lumen; IDA.
DR   GO; GO:0016298; F:lipase activity; IMP.
DR   GO; GO:0006914; P:autophagy; IDA.
DR   GO; GO:0030397; P:membrane degradation; IDA.
DR   GO; GO:0006624; P:vacuolar protein processing/maturation; IDA.
DR   InterPro; IPR002921; Lipase_3.
DR   InterPro; IPR008262; Lipase_AS.
DR   Pfam; PF01764; Lipase_3; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
KW   Hydrolase; Lipid degradation; Autophagy; Endoplasmic reticulum;
KW   Signal-anchor; Transmembrane; Glycoprotein.
FT   DOMAIN        1     14       CYTOPLASMIC.
FT   TRANSMEM     15     35       SIGNAL-ANCHOR (TYPE II MEMBRANE PROTEIN).
FT   DOMAIN       36    520       LUMENAL.
FT   ACT_SITE    332    332       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD    173    173       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    202    202       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    208    208       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   MUTAGEN     332    332       S->A: LOSS OF FUNCTION.
SQ   SEQUENCE   520 AA;  58435 MW;  B56ABAB72B999019 CRC64;

  Query Match             54.7%;  Score 41;  DB 1;  Length 520;
  Best Local Similarity   75.0%;  Pred. No. 56;
  Matches    9;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 GLPAVVGLSPGE 13
              |||||   ||||
Db        346 GLPAVAFESPGE 357


RESULT 7
LOL2_HUMAN
ID   LOL2_HUMAN     STANDARD;      PRT;   774 AA.
AC   Q9Y4K0; Q9BW70; Q9Y5Y8;
```

```
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Lysyl oxidase homolog 2 precursor (EC 1.4.3.-) (Lysyl oxidase-like
DE   protein 2) (Lysyl oxidase related protein 2) (Lysyl oxidase-related
DE   protein WS9-14).
GN   LOXL2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97236759; PubMed=9079631;
RA   Saito H., Papaconstantinou J., Sato H., Goldstein S.;
RT   "Regulation of a novel gene encoding a lysyl oxidase-related protein
RT   in cellular adhesion and senescence.";
RL   J. Biol. Chem. 272:8157-8160(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 202-774 FROM N.A.
RC   TISSUE=Placenta, and Spleen;
RX   MEDLINE=99230328; PubMed=10212285;
RA   Jourdan-Le Saux C., Tronecker H., Bogic L., Bryant-Greenwood G.D.,
RA   Boyd C.D., Csiszar K.;
RT   "The LOXL2 gene encodes a new lysyl oxidase-like protein and is
RT   expressed at high levels in reproductive tissues.";
RL   J. Biol. Chem. 274:12939-12944(1999).
CC   -!- COFACTOR: Copper and LTQ (By similarity).
CC   -!- SUBCELLULAR LOCATION: Extracellular (Potential).
CC   -!- TISSUE SPECIFICITY: Expressed in many tissues. Highest expression
CC       in reproductive tissues, placenta, uterus and prostate.
CC   -!- PTM: The lysine tyrosylquinone cross-link (LTQ) is generated by
CC       condensation of the epsilon-amino group of a lysine with a
CC       topaquinone produced by oxidation of tyrosine.
CC   -!- SIMILARITY: Contains 4 SRCR domains.
CC   -!- SIMILARITY: Belongs to the lysyl oxidase family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U89942; AAB49697.1; -.
DR   EMBL; BC000594; AAH00594.1; -.
DR   EMBL; AF117949; AAD34343.1; -.
DR   Genew; HGNC:6666; LOXL2.
DR   MIM; 606663; -.
DR   GO; GO:0005615; C:extracellular space; TAS.
DR   GO; GO:0005489; F:electron transporter activity; TAS.
DR   GO; GO:0007568; P:aging; TAS.
DR   GO; GO:0007155; P:cell adhesion; TAS.
DR   GO; GO:0006464; P:protein modification; TAS.
DR   InterPro; IPR001695; Lysyl_oxidase.
DR   InterPro; IPR001190; Srcr_receptor.
DR   Pfam; PF01186; Lysyl_oxidase; 1.
DR   Pfam; PF00530; SRCR; 4.
DR   PRINTS; PR00074; LYSYLOXIDASE.
DR   PRINTS; PR00258; SPERACTRCPTR.
DR   ProDom; PD013887; Lysyl_oxidase; 1.
DR   SMART; SM00202; SR; 4.
DR   PROSITE; PS00926; LYSYL_OXIDASE; 1.
DR   PROSITE; PS00420; SRCR_1; 1.
DR   PROSITE; PS50287; SRCR_2; 4.
KW   Oxidoreductase; Copper; Glycoprotein; Repeat; Signal; LTQ.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26    774       LYSYL OXIDASE HOMOLOG 2.
FT   DOMAIN       58    159       SRCR 1.
FT   DOMAIN      188    302       SRCR 2.
FT   DOMAIN      326    425       SRCR 3.
FT   DOMAIN      435    544       SRCR 4.
FT   DOMAIN      548    751       LYSYL-OXIDASE LIKE.
FT   METAL       626    626       COPPER (POTENTIAL).
FT   METAL       628    628       COPPER (POTENTIAL).
FT   METAL       630    630       COPPER (POTENTIAL).
FT   CROSSLNK    653    689       Lysine tyrosylquinone (Lys-Tyr)
FT                                (By similarity).
FT   MOD_RES     689    689       TOPAQUINONE (BY SIMILARITY).
FT   CARBOHYD    288    288       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    455    455       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    644    644       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    295    295       L -> Q (IN REF. 3).
FT   CONFLICT    570    570       M -> L (IN REF. 2).
FT   CONFLICT    652    652       H -> Q (IN REF. 3).
FT   CONFLICT    746    746       C -> S (IN REF. 3).
FT   CONFLICT    774    774       Q -> PVKKPAWSTPVFRPHHIFHGTSPQQLSLNECHVPSP
FT                                SPAPTLSRPLQLCLSSGGKGPSHHSWGAAT (IN REF.
FT                                3).
SQ   SEQUENCE   774 AA;  86724 MW;  9DF5D25D4824BCCD CRC64;

  Query Match             54.7%;  Score 41;  DB 1;  Length 774;
  Best Local Similarity   66.7%;  Pred. No. 83;
  Matches    8;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 GLPAVVGLSPGE 13
              ||||||    ||:
Db        294 GLPAVVSCVPGQ 305


RESULT 8
RISB_PYRFU
ID   RISB_PYRFU     STANDARD;      PRT;   157 AA.
AC   Q8U4L8;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.9) (DMRL synthase)
DE   (Lumazine synthase) (Riboflavin synthase beta chain).
GN   RIBH OR PF0063.
OS   Pyrococcus furiosus.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=2261;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA   Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT   "The complete sequence of the Pyrococcus furiosus genome.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex
```

```
CC       catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-
CC       ribityl-amino-2,4(1H,3H)-pyrimidinedione and L-3,4-dihydrohy-2-
CC       butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit
CC       catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-
CC       amino-2,4(1H,3H)-pyrimidinedione with L-3,4-dihydrohy-2-
CC       butanone-4-phosphate yielding 6,7-dimethyl-8-lumazine (By
CC       similarity).
CC   -!- CATALYTIC ACTIVITY: 2 6,7-dimethyl-8-(1-D-ribityl)lumazine =
CC       riboflavin + 4-(1-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.
CC   -!- PATHWAY: Riboflavin biosynthesis; last step.
CC   -!- SIMILARITY: Belongs to the DMRL synthase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE010132; AAL80187.1; -.
DR   HAMAP; MF_00178; -; 1.
DR   InterPro; IPR002180; DMRL_synthase.
DR   Pfam; PF00885; DMRL_synthase; 1.
DR   ProDom; PD003664; DMRL_synthase; 1.
DR   TIGRFAMs; TIGR00114; ribH; 1.
KW   Riboflavin biosynthesis; Transferase; Complete proteome.
SQ   SEQUENCE   157 AA;  17517 MW;  34BB66E720FD20C1 CRC64;

  Query Match             53.3%;  Score 40;  DB 1;  Length 157;
  Best Local Similarity   42.9%;  Pred. No. 25;
  Matches    6;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 GLPAVVGLSPGEQE 15
              |:| : |: | | |
Db        107 GIPVIFGIVPAEDE 120


RESULT 9
ALF2_PEA
ID   ALF2_PEA       STANDARD;      PRT;   359 AA.
AC   P46257;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Fructose-bisphosphate aldolase, cytoplasmic isozyme 2 (EC 4.1.2.13).
OS   Pisum sativum (Garden pea).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae; Pisum.
OX   NCBI_TaxID=3888;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Leaf;
RA   Pelzer-Reith B., Schnarrenberger C.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate = glycerone
CC       phosphate + D-glyceraldehyde 3-phosphate.
CC   -!- PATHWAY: Glycolysis; sixth step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to class I fructose-bisphosphate aldolase
CC       family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X89829; CAA61947.1; -.
DR   PIR; S58167; S58167.
DR   HSSP; P00883; 1ADO.
DR   InterPro; IPR000741; Aldolase_I.
DR   Pfam; PF00274; glycolytic_enzy; 1.
DR   ProDom; PD001128; Aldolase_I; 1.
DR   PROSITE; PS00158; ALDOLASE_CLASS_I; 1.
KW   Lyase; Schiff base; Glycolysis; Multigene family.
FT   BINDING      52     52       C-1-PHOSPHATE GROUP OF THE SUBSTRATE
FT                                (BY SIMILARITY).
FT   BINDING     143    143       C-1-PHOSPHATE GROUP OF THE SUBSTRATE
FT                                (BY SIMILARITY).
FT   BINDING     226    226       SCHIFF-BASE WITH DIHYDROXYACETONE-P
FT                                (BY SIMILARITY).
FT   ACT_SITE    359    359       ESSENTIAL FOR ENHANCED ACTIVITY OF THE
FT                                ENZYME TOWARD FRUCTOSE 1,6-BISPHOSPHATE
FT                                AS COMPARED WITH FRUCTOSE 1-PHOSPHATE
FT                                (BY SIMILARITY).
SQ   SEQUENCE   359 AA;  38490 MW;  C0CAB16E9CC1B9EF CRC64;

  Query Match             53.3%;  Score 40;  DB 1;  Length 359;
  Best Local Similarity   60.0%;  Pred. No. 56;
  Matches    9;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 AGLPAVVGLSPGEQE 15
              | :|||| || |: |
Db        258 AAVPAVVFLSGGQSE 272


RESULT 10
ALF_CICAR
ID   ALF_CICAR      STANDARD;      PRT;   359 AA.
AC   O65735;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Fructose-bisphosphate aldolase, cytoplasmic isozyme (EC 4.1.2.13).
GN   ALDC.
OS   Cicer arietinum (Chickpea) (Garbanzo).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Cicereae; Cicer.
OX   NCBI_TaxID=3827;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Castellana; TISSUE=Etiolated epicotyl;
RA   Dopico B., Munoz F.J., Labrador E.;
RT   "cDNA and deduced amino-acid sequence of a cytosolic aldolase from
RT   Cicer arietinum L. epicotyls.";
RL   (In) Plant Gene Register PGR98-110.
CC   -!- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate = glycerone
CC       phosphate + D-glyceraldehyde 3-phosphate.
CC   -!- PATHWAY: Glycolysis; sixth step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to class I fructose-bisphosphate aldolase
CC       family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ005041; CAA06308.1; -.
DR   HSSP; P00883; 1ADO.
DR   InterPro; IPR000741; Aldolase_I.
DR   Pfam; PF00274; glycolytic_enzy; 1.
DR   ProDom; PD001128; Aldolase_I; 1.
DR   PROSITE; PS00158; ALDOLASE_CLASS_I; 1.
KW   Lyase; Schiff base; Glycolysis.
FT   BINDING      52     52       C-1-PHOSPHATE GROUP OF THE SUBSTRATE.
FT   BINDING     142    142       C-1-PHOSPHATE GROUP OF THE SUBSTRATE.
FT   BINDING     226    226       SCHIFF-BASE WITH DIHYDROXYACETONE-P.
```

```
FT   ACT_SITE    359    359       ESSENTIAL FOR ENHANCED ACTIVITY OF THE
FT                                ENZYME TOWARD FRUCTOSE 1,6-BISPHOSPHATE
FT                                AS COMPARED WITH FRUCTOSE 1-PHOSPHATE.
SQ   SEQUENCE   359 AA;  38451 MW;  DD68864B745A5195 CRC64;

  Query Match             53.3%;  Score 40;  DB 1;  Length 359;
  Best Local Similarity   60.0%;  Pred. No. 56;
  Matches    9;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 AGLPAVVGLSPGEQE 15
              | :|||| || |: |
Db        258 AAVPAVVFLSGGQSE 272


RESULT 11
VAL1_TMOV
ID   VAL1_TMOV      STANDARD;      PRT;   361 AA.
AC   Q06657;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   AL1 protein.
GN   AL1.
OS   Tomato mottle virus (isolate Florida) (TMoV).
OC   Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX   NCBI_TaxID=36449;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93107858; PubMed=1469361;
RA   Abouzid A.M., Polston J.E., Hiebert E.;
RT   "The nucleotide sequence of tomato mottle virus, a new geminivirus
RT   isolated from tomatoes in Florida.";
RL   J. Gen. Virol. 73:3225-3229(1992).
CC   -!- SIMILARITY: Belongs to the geminiviruses AL1 protein family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L14460; AAC32414.1; -.
DR   PIR; JQ1870; JQ1870.
DR   InterPro; IPR001191; Gemini_AL1.
DR   Pfam; PF00799; Gemini_AL1; 1.
DR   PRINTS; PR00227; GEMCOATAL1.
DR   ProDom; PD000736; Gemini_AL1; 1.
KW   ATP-binding.
FT   NP_BIND     222    229       ATP (BY SIMILARITY).
SQ   SEQUENCE   361 AA;  40516 MW;  8138B65CEEAC6950 CRC64;

  Query Match             53.3%;  Score 40;  DB 1;  Length 361;
  Best Local Similarity   58.3%;  Pred. No. 56;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GLPAVVGLSPGE 13
              |:||:|  :|||
Db        299 GIPAIVLCNPGE 310


RESULT 12
PLB5_SCHPO
ID   PLB5_SCHPO     STANDARD;      PRT;   633 AA.
AC   Q9Y7N6;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Putative lysophospholipase C1450.09c precursor (EC 3.1.1.5)
DE   (Phospholipase B).
GN   SPCC1450.09C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: Catalyzes the release of fatty acids from
CC       lysophospholipids (By similarity).
CC   -!- CATALYTIC ACTIVITY: 2-lysophosphatidylcholine + H(2)O =
CC       glycerophosphocholine + a fatty acid anion.
CC   -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC   -!- SIMILARITY: Belongs to the lysophospholipase family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AL049559; CAB40176.2; -.
DR   GeneDB_SPombe; SPCC1450.09c; -.
DR   InterPro; IPR002642; PLAc.
DR   Pfam; PF01735; PLA2_B; 1.
DR   SMART; SM00022; PLAc; 1.
KW   Hypothetical protein; Lipid degradation; Hydrolase; Glycoprotein;
KW   Signal.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20    633       PUTATIVE LYSOPHOSPHOLIPASE C1450.09C.
FT   CARBOHYD   118    118       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   153    153       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   232    232       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   256    256       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   264    264       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   293    293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   331    331       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   360    360       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   367    367       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   400    400       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   403    403       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   474    474       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   508    508       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   513    513       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD     537     537       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     564     564       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     586     586       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     603     603       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   633 AA;  68292 MW;  49871B2955893D19 CRC64;

  Query Match             53.3%;  Score 40;  DB 1;  Length 633;
  Best Local Similarity   75.0%;  Pred. No. 97;
  Matches    9;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          4 PAVVGLSPGEQE 15
              ||  ||| |||| 
Db         76 PASDGLSTGEQE 87


RESULT 13
YHGF_ECOLI
ID   YHGF_ECOLI     STANDARD;      PRT;   773 AA.
AC   P46837; P76689;
DT   01-NOV-1995 (Rel. 32, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein yhgF.
GN   YHGF OR B3407.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   IDENTIFICATION BY MASS SPECTROMETRY.
RX   MEDLINE=99420866; PubMed=10493123;
RA   Fountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;
RT   "Enrichment of low abundance proteins of Escherichia coli by
RT   hydroxyapatite chromatography.";
RL   Electrophoresis 20:2181-2195(1999).
CC   -!- SIMILARITY: STRONG, TO H.INFLUENZAE HI0568.
CC   -!- SIMILARITY: Contains 1 S1 motif domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U18997; AAA58204.1; ALT_FRAME.
DR   EMBL; U18997; AAA58205.1; ALT_FRAME.
DR   EMBL; AE000416; AAC76432.1; ALT_INIT.
DR   HSSP; P05055; 1SRO.
DR   EcoGene; EG12932; yhgF.
DR   InterPro; IPR008994; Nucleic_acid_OB.
DR   InterPro; IPR003029; S1.
DR   InterPro; IPR006641; YqgFc.
DR   Pfam; PF00575; S1; 1.
DR   SMART; SM00316; S1; 1.
DR   SMART; SM00732; YqgFc; 1.
DR   PROSITE; PS50126; S1; 1.
KW   RNA-binding; Complete proteome.
FT   DOMAIN       651     720       S1 MOTIF.
FT   CONFLICT     754     755       QP -> HA (IN REF. 1; AAA58205).
SQ   SEQUENCE   773 AA;  85119 MW;  EA54D9ED952A8229 CRC64;

  Query Match             53.3%;  Score 40;  DB 1;  Length 773;
  Best Local Similarity   66.7%;  Pred. No. 1.2e+02;
  Matches    8;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 AGLPAVVGLSPG 12
              ||| | :|| ||
Db        324 AGLRATMGLDPG 335


RESULT 14
CADM_MOUSE
ID   CADM_MOUSE     STANDARD;      PRT;   813 AA.
AC   Q9WTP5;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cadherin-22 precursor (PB-cadherin).
GN   CDH22.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGE.
RC   STRAIN=ICR; TISSUE=Brain;
RX   MEDLINE=99326347; PubMed=10398531;
RA   Kitajima K., Koshimizu U., Nakamura T.;
RT   "Expression of a novel type of classic cadherin, PB-cadherin in
RT   developing brain and limb buds.";
RL   Dev. Dyn. 215:206-214(1999).
CC   -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC       They preferentially interact with themselves in a homophilic
CC       manner in connecting cells; cadherins may thus contribute to the
CC       sorting of heterogeneous cell types. PB-cadherins may have a role
CC       in the morphological organization of pituitary gland and brain
CC       tissues.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: Predominantly expressed in brain. Abundant in
CC       olfactory bulb, cerebrum, and cerebellum, less in pons, medulla,
CC       and spinal cord. Low expression in heart. No expression in lung,
CC       liver, spleen, kidney, testis, stomach, intestine, colon, and
CC       placenta.
CC   -!- DEVELOPMENTAL STAGE: Expressed at 9.5 dpc onwards. At 10.5 dpc, in
CC       brain (telencephalic vesicles and isthmus), spinal cord and limb
CC       buds (in the zone of polarizing activity). At 14.5 dpc, in
CC       olfactory bulb and cerebellum.
CC   -!- INDUCTION: Down-regulated by thyroid hormone.
CC   -!- SIMILARITY: Contains 5 cadherin domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB019618; BAA34426.1; -.
DR   MGD; MGI:1341843; Cdh22.
DR   HSSP; P15116; 1NCJ.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000233; Cadherin_C_term.
DR   Pfam; PF00028; cadherin; 5.
DR   Pfam; PF01049; Cadherin_C_term; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 5.
DR   PROSITE; PS00232; CADHERIN_1; 2.
DR   PROSITE; PS50268; CADHERIN_2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW   Signal.
FT   SIGNAL         1     33       POTENTIAL.
FT   CHAIN         34    813       CADHERIN-22.
FT   DOMAIN        33    621       EXTRACELLULAR (POTENTIAL).
```

```
FT   TRANSMEM     622    642       POTENTIAL.
FT   DOMAIN       643    813       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        61    165       CADHERIN 1.
FT   DOMAIN       166    274       CADHERIN 2.
FT   DOMAIN       275    391       CADHERIN 3.
FT   DOMAIN       392    495       CADHERIN 4.
FT   DOMAIN       496    613       CADHERIN 5.
FT   CARBOHYD     159    159       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     463    463       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     609    609       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   813 AA;  88021 MW;  5510F9848D976567 CRC64;

  Query Match             53.3%;  Score 40;  DB 1;  Length 813;
  Best Local Similarity   60.0%;  Pred. No. 1.2e+02;
  Matches    9;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AGLPAVVGLSPGEQE 15
              |  ||   |||| ||
Db         34 ASTPAPSSLSPGAQE 48


RESULT 15
PGCA_BOVIN
ID   PGCA_BOVIN     STANDARD;      PRT;  2364 AA.
AC   P13608; P79117; Q28159;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Aggrecan core protein precursor (Cartilage-specific proteoglycan core
DE   protein) (CSPCP).
GN   AGC1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hering T.M., Kollar J., Huynh T.D.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 563-1056 FROM N.A.
RX   MEDLINE=89380219; PubMed=2528543;
RA   Antonsson P., Heinegaard D., Oldberg A.;
RT   "The keratan sulfate-enriched region of bovine cartilage proteoglycan
RT   consists of a consecutively repeated hexapeptide motif.";
RL   J. Biol. Chem. 264:16170-16173(1989).
RN   [3]
RP   SEQUENCE OF 1609-2113 AND 2151-2364 FROM N.A.
RX   MEDLINE=87270630; PubMed=3111460;
RA   Oldberg A., Antonsson P., Heinegaard D.;
RT   "The partial amino acid sequence of bovine cartilage proteoglycan,
RT   deduced from a cDNA clone, contains numerous Ser-Gly sequences
RT   arranged in homologous repeats.";
RL   Biochem. J. 243:255-259(1987).
RN   [4]
RP   SEQUENCE OF 2114-2150 FROM N.A.
RC   TISSUE=Cartilage;
RX   MEDLINE=93352525; PubMed=8349621;
RA   Fueloep C., Walcz E., Valyon M., Glant T.T.;
RT   "Expression of alternatively spliced epidermal growth factor-like
RT   domains in aggrecans of different species. Evidence for a novel
RT   module.";
RL   J. Biol. Chem. 268:17377-17383(1993).
RN   [5]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=85027710; PubMed=6489519;
RA   Perin J.-P., Bonnet F., Jolles J., Jolles P.;
RT   "Sequence data concerning the protein core of the cartilage
RT   proteoglycan monomers. Characterization of a sequence allowing the
RT   synthesis of an oligonucleotide probe.";
RL   FEBS Lett. 176:37-42(1984).
RN   [6]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=87005253; PubMed=3530809;
RA   Perin J.P., Bonnet F., Jolles P.;
RT   "Structural relationship between link proteins and proteoglycan
RT   monomers.";
RL   FEBS Lett. 206:73-77(1986).
CC   -!- FUNCTION: This proteoglycan is a major component of extracellular
CC       matrix of cartilagenous tissues. A major function of this protein
CC       is to resist compression in cartilage. It binds avidly to
CC       hyaluronic acid via an amino-terminal globular region. May play a
CC       regulatory role in the matrix assembly of the cartilage.
CC   -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
CC       similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=P13608-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P13608-2; Sequence=VSP_003072;
CC   -!- DOMAIN: Two globular domains, G1 and G2, comprise the amino
CC       terminus of the proteoglycan, while another globular region, G3,
CC       makes up the C-terminus. G1 contains link domains and thus
CC       consists of three disulfide-bonded loop structures designated as
CC       the A, B, B' motifs. G2 is similar to G1. The keratan sulfate (KS)
CC       and the chondroitin sulfate (CS) attachment domains lie between G2
CC       and G3.
CC   -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND O-
CC       LINKED (ABOUT 40) OLIGOSACCHARIDES.
CC   -!- PTM: THE KERATAN SULFATE CONTENTS DIFFER CONSIDERABLY BETWEEN
CC       ADULT AND FETAL BOVINE PROTEOGLYCANS.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC   -!- SIMILARITY: Contains 4 link domains.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC   -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U76615; AAB38524.1; -.
DR   EMBL; L07053; -; NOT_ANNOTATED_CDS.
DR   PIR; A34234; A39808.
DR   PIR; T42630; T42630.
DR   HSSP; P08709; 1BF9.
DR   InterPro; IPR002353; AntifreezeII.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001304; Lectin_C.
DR   InterPro; IPR000538; Link.
DR   InterPro; IPR003324; SGXXSG.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF02339; SGXXSG; 61.
DR   Pfam; PF00084; sushi; 1.
DR   Pfam; PF00193; Xlink; 4.
DR   PRINTS; PR00356; ANTIFREEZEII.
DR   PRINTS; PR01265; LINKMODULE.
DR   ProDom; PD000918; Link; 4.
DR   SMART; SM00032; CCP; 1.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00179; EGF_CA; 1.
```

```
DR   SMART; SM00445; LINK; 4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS50026; EGF_3; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
DR   PROSITE; PS00290; IG_MHC; FALSE_NEG.
DR   PROSITE; PS01241; LINK; 4.
KW   Glycoprotein; Proteoglycan; Lectin; Signal; Sushi; EGF-like domain;
KW   Calcium; Alternative splicing; Repeat; Immunoglobulin domain.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17   2364       AGGRECAN CORE PROTEIN.
FT   DOMAIN       25    147       IG-LIKE V-TYPE.
FT   DOMAIN      170    247       LINK 1.
FT   DOMAIN      268    349       LINK 2.
FT   DOMAIN      504    581       LINK 3.
FT   DOMAIN      602    683       LINK 4.
FT   DOMAIN      774    907       23 X 6 AA APPROXIMATE TANDEM REPEATS OF
FT                                E-[EK]-P-F-P-S.
FT   DOMAIN     1433   2112       CS-2.
FT   DOMAIN     2113   2149       EGF-LIKE, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     2114   2364       G3.
FT   DOMAIN     2161   2276       C-TYPE LECTIN.
FT   DOMAIN     2280   2338       SUSHI.
FT   DISULFID     51    133       BY SIMILARITY.
FT   DISULFID    175    246       BY SIMILARITY.
FT   DISULFID    199    220       BY SIMILARITY.
FT   DISULFID    273    348       BY SIMILARITY.
FT   DISULFID    297    318       BY SIMILARITY.
FT   DISULFID    509    580       BY SIMILARITY.
FT   DISULFID    533    554       BY SIMILARITY.
FT   DISULFID    607    682       BY SIMILARITY.
FT   DISULFID    631    652       BY SIMILARITY.
FT   DISULFID   2117   2128       BY SIMILARITY.
FT   DISULFID   2182   2274       BY SIMILARITY.
FT   DISULFID   2250   2266       BY SIMILARITY.
FT   DISULFID   2281   2324       BY SIMILARITY.
FT   DISULFID   2310   2337       BY SIMILARITY.
FT   CARBOHYD    126    126       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    239    239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    333    333       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    387    387       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    611    611       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    667    667       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   2114   2150       Missing (in isoform 2).
FT                                /FTId=VSP_003072.
SQ   SEQUENCE   2364 AA;  246359 MW;  6FF83763420C3D4C CRC64;

  Query Match             53.3%;  Score 40;  DB 1;  Length 2364;
  Best Local Similarity   53.3%;  Pred. No. 3.6e+02;
  Matches    8;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 AGLPAVVGLSPGEQE 15
              :|||   ||  ||:|
Db        971 SGLPVESGLPSGEEE 985


Search completed: May  7, 2004, 12:34:34
Job time : 6.4 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:26:40 ; Search time 28.05 Seconds
(without alignments)
168.726 Million cell updates/sec

Title: US-09-786-214A-15
Perfect score: 75
Sequence: 1 AGLPAVVGLSPGEQE 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	55	73.3	41	11	Q8K408	Q8k408 rattus norv
2	50	66.7	381	16	Q9RRL5	Q9rrl5 deinococcus
3	46	61.3	150	16	Q8PE55	Q8pe55 xanthomonas
4	46	61.3	462	16	Q9KZB0	Q9kzb0 streptomyce
5	44	58.7	335	16	Q82RR7	Q82rr7 streptomyce
6	44	58.7	383	16	Q7V8U3	Q7v8u3 prochloroco
7	44	58.7	399	16	Q8DKG5	Q8dkg5 synechococc
8	43.5	58.0	614	16	Q8Y280	Q8y280 ralstonia s
9	43	57.3	579	16	Q8NRJ7	Q8nrj7 corynebacte
10	43	57.3	582	16	Q8FQM6	Q8fqm6 corynebacte
11	43	57.3	821	17	Q9HPR8	Q9hpr8 halobacteri
12	42.5	56.7	305	16	Q89FB7	Q89fb7 bradyrhizob
13	42	56.0	254	16	Q89BH1	Q89bh1 bradyrhizob
14	42	56.0	326	1	Q9UXP0	Q9uxp0 methanolobu
15	42	56.0	574	2	Q54756	Q54756 synechococc
16	42	56.0	660	9	Q7Y4W9	Q7y4w9 bacteriopha
17	42	56.0	673	13	Q7T2X8	Q7t2x8 gallus gall
18	42	56.0	1541	6	Q8HXL3	Q8hxl3 sus scrofa
19	41	54.7	143	4	Q8NAW2	Q8naw2 homo sapien
20	41	54.7	164	16	Q99QD9	Q99qd9 caulobacter
21	41	54.7	177	17	Q9YAM1	Q9yam1 aeropyrum p
22	41	54.7	214	16	Q9I2S0	Q9i2s0 pseudomonas
23	41	54.7	222	17	Q978I4	Q978i4 thermoplasm
24	41	54.7	239	16	Q89JT2	Q89jt2 bradyrhizob
25	41	54.7	438	17	Q8Q0K1	Q8q0k1 methanosarc
26	41	54.7	448	16	Q8PML2	Q8pml2 xanthomonas
27	41	54.7	520	3	Q8NIL6	Q8nil6 saccharomyc
28	41	54.7	579	4	Q8N158	Q8n158 homo sapien
29	41	54.7	11096	2	Q9L4W3	Q9l4w3 streptomyce
30	40	53.3	85	6	O77786	O77786 canis famil
31	40	53.3	130	17	Q9YFJ5	Q9yfj5 aeropyrum p
32	40	53.3	154	17	Q8ZYJ3	Q8zyj3 pyrobaculum
33	40	53.3	242	16	Q7WCX1	Q7wcx1 bordetella
34	40	53.3	261	16	Q8NQA2	Q8nqa2 corynebacte
35	40	53.3	278	2	Q9F0Y4	Q9f0y4 xanthomonas
36	40	53.3	298	16	Q7W5D6	Q7w5d6 bordetella
37	40	53.3	298	16	Q7VTI2	Q7vti2 bordetella
38	40	53.3	331	16	Q7WEG7	Q7weg7 bordetella
39	40	53.3	331	16	Q7W350	Q7w350 bordetella
40	40	53.3	331	16	Q7W032	Q7w032 bordetella
41	40	53.3	355	10	Q947A7	Q947a7 nitellopsis
42	40	53.3	357	4	Q8NHX1	Q8nhx1 homo sapien
43	40	53.3	358	4	Q9NWD0	Q9nwd0 homo sapien
44	40	53.3	358	10	Q9FUG7	Q9fug7 fragaria an
45	40	53.3	358	12	O91201	O91201 havana toma

ALIGNMENTS

RESULT 1
Q8K408
ID Q8K408 PRELIMINARY; PRT; 41 AA.
AC Q8K408;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Truncated macrophage colony stimulating factor.
GN CSF1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEW.tl;
RX MEDLINE=22069908; PubMed=12074592;
RA Dobbins D.E., Sood R., Hashiramoto A., Hansen C.T., Wilder R.L.,
RA Remmers E.F.;
RT "Mutation of macrophage colony stimulating factor (Csf1) causes
RT osteopetrosis in the tl rat.";
RL Biochem. Biophys. Res. Commun. 294:1114-1120(2002).
DR EMBL; AF514357; AAM54137.1; -.
SQ SEQUENCE 41 AA; 4178 MW; 1D342C19BD18AA41 CRC64;

Query Match 73.3%; Score 55; DB 11; Length 41;
Best Local Similarity 78.6%; Pred. No. 0.1;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GLPAVVGLSPGEQE 15
|||| |||| |||
Db 20 GLPAAAGLSPREQE 33

RESULT 2
Q9RRL5
ID Q9RRL5 PRELIMINARY; PRT; 381 AA.
AC Q9RRL5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

```
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cytochrome P450.
GN   DR2473.
OS   Deinococcus radiodurans.
OC   Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AE002076; AAF12016.1; -.
DR   PIR; F75270; F75270.
DR   TIGR; DR2473; -.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase; Complete proteome.
SQ   SEQUENCE   381 AA;  41940 MW;  F191EA69F1797B53 CRC64;

  Query Match             66.7%;  Score 50;  DB 16;  Length 381;
  Best Local Similarity   100.0%;  Pred. No. 7.4;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLPAVVGLSP 11
              ||||||||||
Db         51 GLPAVVGLSP 60


RESULT 3
Q8PE55
ID   Q8PE55      PRELIMINARY;      PRT;   150 AA.
AC   Q8PE55;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Deoxycytidylate deaminase.
GN   XCC0126.
OS   Xanthomonas campestris (pv. campestris).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xanthomonas.
OX   NCBI_TaxID=340;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33913 / NCPPB 528;
RX   MEDLINE=22022145; PubMed=12024217;
RA   da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA   Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA   Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA   Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA   Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA   Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA   Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA   Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA   Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA   Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA   Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA   Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA   Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA   Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA   Setubal J.C., Kitajima J.P.;
RT   "Comparison of the genomes of two Xanthomonas pathogens with differing
RT   host specificities.";
RL   Nature 417:459-463(2002).
DR   EMBL; AE012107; AAM39445.1; -.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   InterPro; IPR002125; dCMP/cyt_deam.
DR   Pfam; PF00383; dCMP_cyt_deam; 1.
KW   Complete proteome.
SQ   SEQUENCE   150 AA;  16201 MW;  63D9C17D44DC8B43 CRC64;

  Query Match             61.3%;  Score 46;  DB 16;  Length 150;
  Best Local Similarity   60.0%;  Pred. No. 12;
  Matches    9;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 AGLPAVVGLSPGEQE 15
              ||:  || |:||| |
Db        104 AGIKRVVALAPGESE 118


RESULT 4
Q9KZB0
ID   Q9KZB0      PRELIMINARY;      PRT;   462 AA.
AC   Q9KZB0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Putative gntR-family regulator.
GN   SCO4836 OR SC5G8.04.
OS   Streptomyces coelicolor.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Saunders D.C., Harris D.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RX   MEDLINE=21996410; PubMed=12000953;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
CC   -!- SIMILARITY: BELONGS TO THE GNTR FAMILY OF TRANSCRIPTIONAL
CC       REGULATORS.
```

```
DR   EMBL; AL939121; CAB89055.1; -.
DR   GO; GO:0005622; C:intracellular; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR000524; HTH_GntR.
DR   InterPro; IPR000408; Reg_chr_condens.
DR   Pfam; PF00392; gntR; 1.
DR   PRINTS; PR00035; HTHGNTR.
DR   SMART; SM00345; HTH_GNTR; 1.
DR   PROSITE; PS00626; RCC1_2; 1.
KW   DNA-binding; Transcription regulation; Complete proteome.
SQ   SEQUENCE   462 AA;  48831 MW;  DF0ECDA651E13317 CRC64;

  Query Match             61.3%;  Score 46;  DB 16;  Length 462;
  Best Local Similarity   60.0%;  Pred. No. 41;
  Matches    9;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 AGLPAVVGLSPGEQE 15
              ||| ||:|| || ::
Db        392 AGLHAVLGLPPGTEQ 406


RESULT 5
Q82RR7
ID   Q82RR7      PRELIMINARY;      PRT;   335 AA.
AC   Q82RR7;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative terpene cyclase.
GN   TPC1 OR SAV76.
OS   Streptomyces avermitilis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX   MEDLINE=22608306; PubMed=12692562;
RA   Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA   Sakaki Y., Hattori M., Omura S.;
RT   "Complete genome sequence and comparative analysis of the industrial
RT   microorganism Streptomyces avermitilis.";
RL   Nat. Biotechnol. 21:526-531(2003).
DR   EMBL; AP005021; BAC67785.1; -.
DR   InterPro; IPR008949; Terpenoid_synth.
KW   Complete proteome.
SQ   SEQUENCE   335 AA;  36480 MW;  49B8477E2D52666F CRC64;

  Query Match             58.7%;  Score 44;  DB 16;  Length 335;
  Best Local Similarity   72.7%;  Pred. No. 62;
  Matches    8;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GLPAVVGLSPG 12
              ||||  |:|||
Db          8 GLPAPAGISPG 18


RESULT 6
Q7V8U3
ID   Q7V8U3      PRELIMINARY;      PRT;   383 AA.
AC   Q7V8U3;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phage integrase.
GN   PMT0234.
OS   Prochlorococcus marinus (strain MIT 9313).
OC   Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC   Prochlorococcus.
OX   NCBI_TaxID=74547;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22825698; PubMed=12917642;
RA   Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,
RA   Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,
RA   Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M.,
RA   Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,
RA   Webb E.A., Zinser E.R., Chisholm S.W.;
RT   "Genome divergence in two Prochlorococcus ecotypes reflects oceanic
RT   niche differentiation.";
RL   Nature 424:1042-1047(2003).
DR   EMBL; BX572095; CAE20409.1; -.
KW   Complete proteome.
SQ   SEQUENCE   383 AA;  43234 MW;  CF956FE676F2E59B CRC64;

  Query Match             58.7%;  Score 44;  DB 16;  Length 383;
  Best Local Similarity   75.0%;  Pred. No. 71;
  Matches    9;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 GLPAVVGLSPGE 13
              |||| | |:|||
Db        193 GLPASVRLTPGE 204


RESULT 7
Q8DKG5
ID   Q8DKG5      PRELIMINARY;      PRT;   399 AA.
AC   Q8DKG5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Tlr0894 protein.
GN   TLR0894.
OS   Synechococcus elongatus (Thermosynechococcus elongatus).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX   NCBI_TaxID=32046;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BP-1;
RX   MEDLINE=22225144; PubMed=12240834;
RA   Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA   Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA   Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA   Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the thermophilic cyanobacterium
RT   Thermosynechococcus elongatus BP-1.";
RL   DNA Res. 9:123-130(2002).
DR   EMBL; AP005372; BAC08446.1; -.
DR   InterPro; IPR000437; Prok_lipoprot_S.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Complete proteome.
SQ   SEQUENCE   399 AA;  45259 MW;  61F60CE6B22A3F94 CRC64;

  Query Match             58.7%;  Score 44;  DB 16;  Length 399;
  Best Local Similarity   64.3%;  Pred. No. 74;
  Matches    9;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 GLPAVVGLSPGEQE 15
              ||| | |:|| ||:
Db        273 GLPFVKGVSPYEQQ 286


RESULT 8
Q8Y280
```

```
ID   Q8Y280      PRELIMINARY;      PRT;   614 AA.
AC   Q8Y280;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Probable ATP-binding transport ABC transporter protein.
GN   RSC0456 OR RS04444.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Burkholderiaceae; Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646059; CAD13984.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR   GO; GO:0000166; F:nucleotide binding; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR001623; DnaJ_N.
DR   Pfam; PF00005; ABC_tran; 1.
DR   ProDom; PD000006; ABC_transporter; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR   PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
DR   PROSITE; PS00636; DNAJ_1; 1.
KW   Complete proteome.
SQ   SEQUENCE   614 AA;  69240 MW;  E293355B85872142 CRC64;

  Query Match             58.0%;  Score 43.5;  DB 16;  Length 614;
  Best Local Similarity   45.8%;  Pred. No. 1.4e+02;
  Matches   11;  Conservative    2;  Mismatches    0;  Indels   11;  Gaps    1;

Qy          3 LPAVVG-----------LSPGEQE 15
              |||:||           |||||||:
Db        501 LPALVGLLDEVSNWSLRLSPGEQQ 524


RESULT 9
Q8NRJ7
ID   Q8NRJ7      PRELIMINARY;      PRT;   579 AA.
AC   Q8NRJ7;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Sulfate permease and related transporters (MFS superfamily).
GN   CGL1051.
OS   Corynebacterium glutamicum (Brevibacterium flavum).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX   NCBI_TaxID=1718;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA   Nakagawa S.;
RT   "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP005277; BAB98444.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0008271; F:sulfate porter activity; IEA.
DR   GO; GO:0008272; P:sulfate transport; IEA.
DR   InterPro; IPR006162; Ppantne_S.
DR   InterPro; IPR002645; STAS.
DR   InterPro; IPR001902; Sulph_transpt.
DR   Pfam; PF01740; STAS; 1.
DR   Pfam; PF00916; Sulfate_transp; 1.
DR   TIGRFAMs; TIGR00815; sulP; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS50801; STAS; 1.
KW   Complete proteome.
SQ   SEQUENCE   579 AA;  61660 MW;  B70EA8BBE0C4C100 CRC64;

  Query Match             57.3%;  Score 43;  DB 16;  Length 579;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGLPAVVGL 9
              |||||||||
Db         48 AGLPAVVGL 56


RESULT 10
Q8FQM6
ID   Q8FQM6      PRELIMINARY;      PRT;   582 AA.
AC   Q8FQM6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Putative sulfate transport protein.
GN   CE1093.
OS   Corynebacterium efficiens.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX   NCBI_TaxID=152794;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA   Kawarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA   Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA   Usuda Y., Sugimoto S.;
RT   "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP005217; BAC17903.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0008271; F:sulfate porter activity; IEA.
DR   GO; GO:0008272; P:sulfate transport; IEA.
DR   InterPro; IPR002645; STAS.
DR   InterPro; IPR001902; Sulph_transpt.
DR   Pfam; PF01740; STAS; 1.
DR   Pfam; PF00916; Sulfate_transp; 1.
DR   TIGRFAMs; TIGR00815; sulP; 1.
DR   PROSITE; PS50801; STAS; 1.
KW   Complete proteome.
SQ   SEQUENCE   582 AA;  62545 MW;  CA67CFCAEB04AE58 CRC64;

  Query Match             57.3%;  Score 43;  DB 16;  Length 582;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGLPAVVGL 9
              |||||||||
Db         51 AGLPAVVGL 59


RESULT 11
Q9HPR8
ID   Q9HPR8      PRELIMINARY;      PRT;   821 AA.
AC   Q9HPR8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   DNA helicase.
GN   HEL OR VNG1501G.
OS   Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
```

```
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR   EMBL; AE005064; AAG19799.1; -.
DR   PIR; C84304; C84304.
DR   GO; GO:0004386; F:helicase activity; IEA.
DR   GO; GO:0000166; F:nucleotide binding; IEA.
DR   GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001478; PDZ.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00228; PDZ; 1.
KW   Helicase; Complete proteome.
SQ   SEQUENCE   821 AA;  89848 MW;  C454C76B984A5702 CRC64;

  Query Match           57.3%;  Score 43;  DB 17;  Length 821;
  Best Local Similarity  90.0%;  Pred. No. 2.3e+02;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          5 AVVGLSPGEQ 14
              |||||| ||
Db        326 AVVGLSPAEQ 335


RESULT 12
Q89FB7
ID   Q89FB7      PRELIMINARY;      PRT;   305 AA.
AC   Q89FB7;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Dihydrodipicolinate synthase.
GN   DAPA OR BLR6784.
OS   Bradyrhizobium japonicum.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Bradyrhizobiaceae; Bradyrhizobium.
OX   NCBI_TaxID=375;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=USDA 110;
RX   MEDLINE=22484998; PubMed=12597275;
RA   Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA   Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA   Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA   Tabata S.;
RT   "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT   Bradyrhizobium japonicum USDA110.";
RL   DNA Res. 9:189-197(2002).
DR   EMBL; AP005959; BAC52049.1; -.
DR   InterPro; IPR002220; DHDPS.
DR   Pfam; PF00701; DHDPS; 1.
DR   PRINTS; PR00146; DHPICSNTHASE.
KW   Complete proteome.
SQ   SEQUENCE   305 AA;  32247 MW;  EB02891D5EE5A4D8 CRC64;

  Query Match           56.7%;  Score 42.5;  DB 16;  Length 305;
  Best Local Similarity  45.8%;  Pred. No. 98;
  Matches   11;  Conservative    2;  Mismatches    2;  Indels    9;  Gaps    1;

Qy          1 AGLPAVV---------GLSPGEQE 15
                  ||:||:|         ||| |||
Db         41 AGVPAIVVCETAGEAPTLSPAEQE 64


RESULT 13
Q89BH1
ID   Q89BH1      PRELIMINARY;      PRT;   254 AA.
AC   Q89BH1;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Bll8182 protein.
GN   BLL8182.
OS   Bradyrhizobium japonicum.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Bradyrhizobiaceae; Bradyrhizobium.
OX   NCBI_TaxID=375;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=USDA 110;
RX   MEDLINE=22484998; PubMed=12597275;
RA   Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA   Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA   Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA   Tabata S.;
RT   "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT   Bradyrhizobium japonicum USDA110.";
RL   DNA Res. 9:189-197(2002).
DR   EMBL; AP005964; BAC53447.1; -.
DR   InterPro; IPR002559; Transposase_11.
DR   Pfam; PF01609; Transposase_11; 1.
KW   Complete proteome.
SQ   SEQUENCE   254 AA;  29160 MW;  76C1ED8F28D5C758 CRC64;

  Query Match           56.0%;  Score 42;  DB 16;  Length 254;
  Best Local Similarity  66.7%;  Pred. No. 98;
  Matches    8;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 GLPAVVGLSPGE 13
              |||  : |||||
Db        131 GLPVQLALSPGE 142


RESULT 14
Q9UXP0
ID   Q9UXP0      PRELIMINARY;      PRT;   326 AA.
AC   Q9UXP0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   F420-dependent N5,N10-methylene-tetrahydromethanopterin reductase,
DE   putative.
GN   FFDA.
OS   Methanolobus tindarius.
OC   Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC   Methanosarcinales; Methanosarcinaceae; Methanolobus.
OX   NCBI_TaxID=2221;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 2278;
RA   Westenberg D.J., Braune A., Ruppert C., Mueller V., Herzberg C.,
RA   Gottschalk G., Blaut M.;
RT   "The F420H2-dehydrogenase from Methanolobus tindarius: Cloning of the
RT   ffd operon and expression of the genes in Escherichia coli.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ011519; CAB56639.1; -.
DR   PIR; T45226; T45226.
DR   InterPro; IPR002103; Bac_luciferase.
DR   Pfam; PF00296; bac_luciferase; 1.
SQ   SEQUENCE   326 AA;  34043 MW;  16F3AB9733A45D82 CRC64;

  Query Match           56.0%;  Score 42;  DB 1;  Length 326;
  Best Local Similarity  57.1%;  Pred. No. 1.3e+02;
```

```
  Matches    8;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 AGLPAVVGLSPGEQ 14
              :|  |::|| ||||
Db         84 SGGRAILGLGPGEQ 97


RESULT 15
Q54756
ID   Q54756      PRELIMINARY;      PRT;   574 AA.
AC   Q54756;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to plant sulfate transporter.
OS   Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX   NCBI_TaxID=1140;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC 7942;
RA   Phung L.T., Haselkorn R.;
RT   "Gene encoding biotin carboxylase subunit of acetyl-CoA carboxylase
RT   from cyanobacterium Synechococcus sp. PCC 7942.";
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U59234; AAB88215.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR   GO; GO:0008271; F:sulfate porter activity; IEA.
DR   GO; GO:0008272; P:sulfate transport; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR002645; STAS.
DR   InterPro; IPR001902; Sulph_transpt.
DR   Pfam; PF01740; STAS; 1.
DR   Pfam; PF00916; Sulfate_transp; 1.
DR   TIGRFAMs; TIGR00815; sulP; 1.
DR   PROSITE; PS50801; STAS; 1.
SQ   SEQUENCE   574 AA;  62328 MW;  7BDECD65C67BBB00 CRC64;

  Query Match             56.0%;  Score 42;  DB 2;  Length 574;
  Best Local Similarity   88.9%;  Pred. No. 2.3e+02;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGLPAVVGL 9
              |||||:|||
Db         50 AGLPAIVGL 58


Search completed: May  7, 2004, 12:37:57
Job time : 29.2167 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:19:40 ; Search time 41.85 Seconds
(without alignments)
101.272 Million cell updates/sec

Title: US-09-786-214A-15
Perfect score: 75
Sequence: 1 AGLPAVVGLSPGEQE 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	75	100.0	15	3	AAY84269	Aay84269 Peptide d
2	75	100.0	20	3	AAY84265	Aay84265 Truncated
3	75	100.0	25	3	AAY84264	Aay84264 Peptide o
4	65	86.7	13	3	AAY84268	Aay84268 Peptide d
5	65	86.7	14	3	AAY84266	Aay84266 Peptide d
6	61	81.3	13	3	AAY84267	Aay84267 Peptide d
7	45	60.0	234	4	AAB36208	Aab36208 Human imm
8	44	58.7	116	4	ABG24038	Abg24038 Novel hum
9	43	57.3	473	4	AAU40252	Aau40252 Propionib
10	43	57.3	473	6	ABM36771	Abm36771 Propionib
11	43	57.3	565	4	AAB76817	Aab76817 Corynebac
12	43	57.3	579	4	AAG90917	Aag90917 C glutami
13	42	56.0	470	6	AAE34724	Aae34724 Streptomy
14	42	56.0	475	6	AAE34732	Aae34732 Streptomy
15	42	56.0	475	6	AAE34729	Aae34729 Streptomy
16	41	54.7	106	5	ABJ10397	Abj10397 Mutant an
17	41	54.7	106	5	ABJ10395	Abj10395 Mutant an
18	41	54.7	143	5	AAO21677	Aao21677 Human sec
19	41	54.7	143	7	ADB64633	Adb64633 Human pro
20	41	54.7	154	4	AAM25921	Aam25921 Human pro
21	41	54.7	241	5	AAM48925	Aam48925 scFv anti
22	41	54.7	294	1	AAP82484	Aap82484 Tropoelas
23	41	54.7	370	2	AAY14924	Aay14924 Amino aci
24	41	54.7	370	6	ABP70895	Abp70895 Mycobacte
25	41	54.7	372	5	ABB73530	Abb73530 M vaccae
26	41	54.7	396	3	AAB51784	Aab51784 Gene 15 h
27	41	54.7	530	7	ADD49105	Add49105 Human NOV
28	41	54.7	549	7	ADD49091	Add49091 Human NOV
29	41	54.7	579	5	ABG70277	Abg70277 Human Gly
30	41	54.7	579	5	ABG97356	Abg97356 Human CGD
31	41	54.7	579	6	ABR39111	Abr39111 Human GPC
32	41	54.7	579	7	ADD49087	Add49087 Human NOV
33	41	54.7	579	7	ADD49107	Add49107 Human NOV
34	41	54.7	579	7	ADD49089	Add49089 Human NOV
35	41	54.7	592	7	ADD49099	Add49099 Human NOV
36	41	54.7	774	3	AAB00077	Aab00077 Human lys
37	41	54.7	774	5	ABB07649	Abb07649 Human LOR
38	41	54.7	774	6	ABU03509	Abu03509 Angiogene
39	41	54.7	774	6	ABU57639	Abu57639 Different
40	41	54.7	774	6	ABR41070	Abr41070 Human MAP
41	41	54.7	11096	4	AAE10129	Aae10129 Streptomy
42	40	53.3	119	4	AAU51147	Aau51147 Propionib
43	40	53.3	119	6	ABM47666	Abm47666 Propionib
44	40	53.3	132	3	AAG36582	Aag36582 Arabidops
45	40	53.3	215	5	ABB97875	Abb97875 Human sec

ALIGNMENTS

RESULT 1
AAY84269
ID AAY84269 standard; peptide; 15 AA.
XX
AC AAY84269;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Example 2; Page 40; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying

CC functional variants and mimetics
XX
SQ Sequence 15 AA;

Query Match 100.0%; Score 75; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 4.7e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLPAVVGLSPGEQE 15
||||||||||||||||
Db 1 AGLPAVVGLSPGEQE 15

RESULT 2
AAY84265
ID AAY84265 standard; peptide; 20 AA.
XX
AC AAY84265;
XX
DT 12-JUL-2000 (first entry)
XX
DE Truncated macrophage colony stimulating factor tumour antigen.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
DR N-PSDB; AAZ99675.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Claim 3; Page 64; 74pp; English.
XX
CC The present sequence represents a truncated tumour rejection antigen
CC precursor, and is encoded by a truncated alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics
XX
SQ Sequence 20 AA;

Query Match 100.0%; Score 75; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLPAVVGLSPGEQE 15
||||||||||||||||
Db 2 AGLPAVVGLSPGEQE 16

RESULT 3
AAY84264
ID AAY84264 standard; peptide; 25 AA.
XX
AC AAY84264;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide of alternate reading frame of macrophage colony stimulating gene.
XX
KW Renal cell carcinoma; antigen; cytotoxic T lymphocyte;
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
DR N-PSDB; AAZ99672.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Claim 1; Page 64; 74pp; English.
XX
CC The present sequence represents a tumour rejection antigen precursor, and
CC is encoded by an alternative open reading frame (ORF) of human macrophage
CC colony stimulating gene. Peptides derived from the alternative ORF of
CC macrophage-colony stimulating factor, when presented by an antigen
CC presenting cell having a human leukocyte antigen (HLA) class I molecule,
CC effectively induce the activation and proliferation of CD8+ cytotoxic T
CC lymphocytes. Polypeptide and nucleic acids derived from the alternate ORF
CC of macrophage-colony stimulating factor are useful for enriching
CC selectively a population of T lymphocytes with CD8+ T lymphocytes. They
CC are also useful for diagnosing a disorder characterized by expression of
CC the polypeptide, and for identifying functional variants and mimetics
XX
SQ Sequence 25 AA;

Query Match 100.0%; Score 75; DB 3; Length 25;
Best Local Similarity 100.0%; Pred. No. 8.2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLPAVVGLSPGEQE 15
||||||||||||||||
Db 2 AGLPAVVGLSPGEQE 16

RESULT 4
AAY84268
ID AAY84268 standard; peptide; 13 AA.
XX
AC AAY84268;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;

KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Example 2; Page 40; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics
XX
SQ Sequence 13 AA;

Query Match 86.7%; Score 65; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 LPAVVGLSPGEQE 15
|||||||||||||
Db 1 LPAVVGLSPGEQE 13

RESULT 5
AAY84266
ID AAY84266 standard; peptide; 14 AA.
XX
AC AAY84266;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
PS Claim 2; Page 39; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics
XX
SQ Sequence 14 AA;

Query Match 86.7%; Score 65; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 LPAVVGLSPGEQE 15
|||||||||||||
Db 1 LPAVVGLSPGEQE 13

RESULT 6
AAY84267
ID AAY84267 standard; peptide; 13 AA.
XX
AC AAY84267;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX

PS Example 2; Page 40; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics
XX
SQ Sequence 13 AA;

Query Match 81.3%; Score 61; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 PAVVGLSPGEQE 15
||||||||||||
Db 1 PAVVGLSPGEQE 12

RESULT 7
AAB36208
ID AAB36208 standard; protein; 234 AA.
XX
AC AAB36208;
XX
DT 15-FEB-2001 (first entry)
XX
DE Human immune system associated protein HISAP-6.
XX
KW Human; immune system associated protein; HISAP-6; immune disorder;
KW infection; autoimmune disease; cancer.
XX
OS Homo sapiens.
XX
PN US6135941-A.
XX
PD 24-OCT-2000.
XX
PF 27-MAR-1998; 98US-00049672.
XX
PR 27-MAR-1998; 98US-00049672.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Yue H, Lal P, Corley NC, Guegler KJ, Baughn MR;
PI Hillman JL, Au-Young J;
XX
DR WPI; 2001-030926/04.
DR N-PSDB; AAC66524.
XX
PT New human immune system associated proteins (HISAP) and polynucleotides
PT encoding the HISAP, useful for diagnosing, treating or preventing immune
PT or cell proliferative disorders or infections.
XX
PS Claim 1; Col 59-60; 54pp; English.
XX
CC The present invention provides the coding and protein sequences for a
CC number of human immune system associated proteins (HISAPs). These can be
CC used in the diagnosis and treatment of various autoimmune disorders,
CC infections and cell proliferation diseases. The diseases include AIDS,
CC adult respiratory distress syndrome, anaemia, asthma, atherosclerosis,
CC Crohn's disease, irritable bowel syndrome, multiple sclerosis, myasthenia
CC gravis, osteoarthritis, rheumatoid arthritis, scleroderma, systemic lupus
CC erythematosus, arteriosclerosis, cirrhosis and cancer
XX
SQ Sequence 234 AA;

Query Match 60.0%; Score 45; DB 4; Length 234;
Best Local Similarity 72.7%; Pred. No. 52;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 4 PAVVGLSPGEQ 14
||:| |||||:
Db 28 PAIVSLSPGER 38

RESULT 8
ABG24038
ID ABG24038 standard; protein; 116 AA.
XX
AC ABG24038;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #24029.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS88225.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 54397; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 116 AA;

Query Match 58.7%; Score 44; DB 4; Length 116;

```
  Best Local Similarity   64.3%;  Pred. No. 35;
  Matches    9;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 GLPAVVGLSPGEQE 15
              ||| : |||| |:|
Db         90 GLPXLRGLSPVERE 103


RESULT 9
AAU40252
ID   AAU40252 standard; protein; 473 AA.
XX
AC   AAU40252;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Propionibacterium acnes immunogenic protein #1148.
XX
KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
XX
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2.
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US012865.
XX
PR   21-APR-2000; 2000US-0199047P.
PR   02-JUN-2000; 2000US-0208841P.
PR   07-JUL-2000; 2000US-0216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW,  Persing DH,  Mitcham JL,  Wang SS,  Bhatia A;
PI   L'maisonneuve J,  Zhang Y,  Jen S,  Carter D;
XX
DR   WPI; 2001-616774/71.
DR   N-PSDB; AAS59511.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris.
XX
PS   Example 1; SEQ ID NO 1447; 1069pp; English.
XX
CC   Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
CC   and determining the amount of bound protein in the sample. The
CC   polypeptides may be used as antigens in the production of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
CC   downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC   this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 473 AA;

  Query Match           57.3%;  Score 43;  DB 4;  Length 473;

  Best Local Similarity   63.6%;  Pred. No. 2.3e+02;
  Matches    7;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 GLPAVVGLSPG 12
              ||||::||: |
Db        446 GLPAIIGLAAG 456


RESULT 10
ABM36771
ID   ABM36771 standard; protein; 473 AA.
XX
AC   ABM36771;
XX
DT   20-OCT-2003  (first entry)
XX
DE   Propionibacterium acnes predicted ORF-encoded polypeptide #1447.
XX
KW   Acne vulgaris; antiseborrhoeic; dermatological; antibacterial;
KW   immunostimulant; immune response; vaccine.
XX
OS   Propionibacterium acnes.
XX
PN   WO2003033515-A1.
XX
PD   24-APR-2003.
XX
PF   11-OCT-2002; 2002WO-US032727.
XX
PR   15-OCT-2001; 2001US-00978825.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Mitcham JL,  Skeiky YAW,  Persing DH,  Bhatia A,  Maisonneuve JL;
PI   Zhang Y,  Wang S,  Jen S,  Lodes MJ,  Benson DR,  Jones R,  Carter D;
PI   Barth B,  Vallieve-Douglass J;
XX
DR   WPI; 2003-381789/36.
DR   N-PSDB; ACF64440.
XX
PT   New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT   polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT   or for stimulating an immune response specific for a P. acnes protein.
XX
PS   Example 1; SEQ ID NO 1447; 1481pp; English.
XX
CC   The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC   encoding a Propionibacterium acnes protein. The invention also relates to
CC   polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC   immunogenic fragments of P. acnes polypeptides. The invention
CC   additionally encompasses expression vectors and host cells comprising a
CC   polynucleotide of the invention; antibodies against polypeptides of the
CC   invention; fusion proteins comprising a polypeptide of the invention; a
CC   method for stimulating an immune response specific for a P. acnes
CC   polypeptide and an isolated T cell population comprising T cells prepared
CC   via this method; a vaccine composition (comprising P. acnes polypeptides,
CC   polynucleotides, antibodies, fusion proteins, T cell populations, or
CC   antigen-presenting cells that express the polypeptide); a method and kit
CC   for detecting or determining the presence or absence of P. acnes in a
CC   patient; and a method for inhibiting the development of P. acnes in a
CC   patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC   proteins, T cell populations or antigen-presenting cells that express the
CC   polypeptides are useful for diagnosing, preventing or treating acne
CC   vulgaris, or for stimulating an immune response specific for a P. acnes
CC   protein. The polynucleotides can also be used as probes or primers for
CC   nucleic acid hybridisation. The vaccine composition is useful for the
CC   stimulation of an immune response against P. acnes, or for treating acne,
CC   and the kit is useful for performing a diagnostic assay. The present
CC   sequence represents a polypeptide predicted to be encoded by an ORF (open
CC   reading frame) contained within the P. acnes polynucleotides of the
CC   invention. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
```

```
XX
SQ   Sequence 473 AA;

  Query Match             57.3%;  Score 43;  DB 6;  Length 473;
  Best Local Similarity   63.6%;  Pred. No. 2.3e+02;
  Matches    7;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 GLPAVVGLSPG 12
              ||||::||: |
Db        446 GLPAIIGLAAG 456


RESULT 11
AAB76817
ID   AAB76817 standard; protein; 565 AA.
XX
AC   AAB76817;
XX
DT   11-APR-2001  (first entry)
XX
DE   Corynebacterium glutamicum MCT protein SEQ ID NO:616.
XX
KW   Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
KW   membrane construction and membrane transport protein; petroleum spill;
KW   hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW   identification; microorganism; fine chemical production; transformation;
KW   genome mapping; genetic engineering.
XX
OS   Corynebacterium glutamicum.
XX
PN   WO200100805-A2.
XX
PD   04-JAN-2001.
XX
PF   23-JUN-2000; 2000WO-IB000926.
XX
PR   25-JUN-1999;   99US-0141031P.
PR   08-JUL-1999;   99DE-01031454.
PR   08-JUL-1999;   99DE-01031478.
PR   08-JUL-1999;   99DE-01031563.
PR   09-JUL-1999;   99DE-01032122.
PR   09-JUL-1999;   99DE-01032124.
PR   09-JUL-1999;   99DE-01032125.
PR   09-JUL-1999;   99DE-01032128.
PR   09-JUL-1999;   99DE-01032180.
PR   09-JUL-1999;   99DE-01032182.
PR   09-JUL-1999;   99DE-01032190.
PR   09-JUL-1999;   99DE-01032191.
PR   09-JUL-1999;   99DE-01032209.
PR   09-JUL-1999;   99DE-01032212.
PR   09-JUL-1999;   99DE-01032227.
PR   09-JUL-1999;   99DE-01032228.
PR   09-JUL-1999;   99DE-01032229.
PR   09-JUL-1999;   99DE-01032230.
PR   14-JUL-1999;   99DE-01032927.
PR   14-JUL-1999;   99DE-01033005.
PR   14-JUL-1999;   99DE-01033006.
PR   27-AUG-1999;   99DE-01040764.
PR   27-AUG-1999;   99DE-01040765.
PR   27-AUG-1999;   99DE-01040766.
PR   27-AUG-1999;   99DE-01040830.
PR   27-AUG-1999;   99DE-01040831.
PR   27-AUG-1999;   99DE-01040832.
PR   27-AUG-1999;   99DE-01040833.
PR   31-AUG-1999;   99DE-01041378.
PR   31-AUG-1999;   99DE-01041379.
PR   31-AUG-1999;   99DE-01041395.
PR   03-SEP-1999;   99DE-01042077.
PR   03-SEP-1999;   99DE-01042078.
PR   03-SEP-1999;   99DE-01042079.
PR   03-SEP-1999;   99DE-01042088.
XX
PA   (BADI ) BASF AG.
XX
PI   Pompejus M,  Kroeger B,  Schroeder H,  Zelder O,  Haberhauer G;
XX
DR   WPI; 2001-071486/08.
DR   N-PSDB; AAF68050.
XX
PT   Corynebacteium glutamicum nucleic acids encoding membrane construction
PT   and membrane transport proteins or their portions, useful for typing or
PT   identifying C. glutamicum or related bacteria, and as markers for
PT   transformation.
XX
PS   Claim 20; Page 1030-1031; 1119pp; English.
XX
CC   AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
CC   construction and membrane transport (MCT) proteins given in AAB76510 to
CC   AAB76847. The MCT nucleic acids and proteins are useful in the
CC   identification of microorganisms which can be used to produce fine
CC   chemicals, for modulating fine chemical production in C. glutamicum or
CC   related bacteria (e.g. Brevibacterium lactofermentum), the typing or
CC   identification of C. glutamicum or related bacteria, as reference points
CC   for mapping C. glutamicum genome, and as markers for transformation.
CC   AAF68082 and AAF68082 represent sequencing primers which are used in an
CC   example from the present invention
XX
SQ   Sequence 565 AA;

  Query Match             57.3%;  Score 43;  DB 4;  Length 565;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGLPAVVGL 9
              |||||||||
Db         34 AGLPAVVGL 42


RESULT 12
AAG90917
ID   AAG90917 standard; protein; 579 AA.
XX
AC   AAG90917;
XX
DT   26-SEP-2001  (first entry)
XX
DE   C glutamicum protein fragment SEQ ID NO: 4671.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001.
XX
PF   18-DEC-2000; 2000EP-00127688.
XX
PR   16-DEC-1999;   99JP-00377484.
PR   07-APR-2000; 2000JP-00159162.
PR   03-AUG-2000; 2000JP-00280988.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Nakagawa S,  Mizoguchi H,  Ando S,  Hayashi M,  Ochiai K,  Yokoi H;
PI   Tateishi N,  Senoh A,  Ikeda M,  Ozaki A;
XX
DR   WPI; 2001-376931/40.
DR   N-PSDB; AAH66136.
XX
PT   Novel polynucleotides derived from Coryneform bacteria, for identifying
PT   mutation point of a gene, measuring expression of a gene, analyzing
PT   expression profile or pattern of a gene and identifying homologous gene.
XX
PS   Claim 17; SEQ ID NO 4671; 246pp + Sequence Listing; English.
```

XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and analysing
CC the expression profile or expression pattern of a gene derived from
CC Coryneform bacterium, and identifying a homologue of a gene derived from
CC coryneform bacterium. Coryneform bacteria are useful for producing amino
CC acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described in the
CC exemplification of the invention. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from the European Patent Office
XX
SQ Sequence 579 AA;

Query Match 57.3%; Score 43; DB 4; Length 579;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGLPAVVGL 9
||||||||| |
Db 48 AGLPAVVGL 56

RESULT 13
AAE34724
ID AAE34724 standard; protein; 470 AA.
XX
AC AAE34724;
XX
DT 14-MAY-2003 (first entry)
XX
DE Streptomyces rimosus ema3 protein.
XX
KW P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;
KW emamectin; insecticide; ema3 protein.
XX
OS Streptomyces rimosus.
XX
PN WO200292801-A2.
XX
PD 21-NOV-2002.
XX
PF 15-MAY-2002; 2002WO-EP005363.
XX
PR 16-MAY-2001; 2001US-0291149P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
PI Buckel TG;
XX
DR WPI; 2003-140280/13.
DR N-PSDB; AAD53019.
XX
PT Novel polypeptide that exhibits enzymatic activity of P450 monooxygenase
PT and capable of regioselectively oxidizing alcohol of avermectin useful
PT for making emamectin from avermectin.
XX
PS Claim 17; Page 107-108; 157pp; English.
XX
CC The present invention relates to novel proteins that exhibit an enzymatic
CC activity of P450 monooxygenase and capable of regioselectively oxidising
CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
CC The invention also relates to ferrodoxins and ferrodoxin reductases that
CC are active with the P450 monooxygenases. Sequences of the invention are
CC useful for producing 4-keto-avermectin from avermectin, which is useful
CC for producing emamectin. Emamectin is useful as an insecticide. The
CC present sequence is Streptomyces rimosus ema3 protein
XX
SQ Sequence 470 AA;

Query Match 56.0%; Score 42; DB 6; Length 470;
Best Local Similarity 72.7%; Pred. No. 3.3e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 LPAVVGLSPGE 13
||: ||| |||
Db 44 LPSYVGLHPGE 54

RESULT 14
AAE34732
ID AAE34732 standard; protein; 475 AA.
XX
AC AAE34732;
XX
DT 14-MAY-2003 (first entry)
XX
DE Streptomyces rimosus emal1 protein.
XX
KW P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;
KW emamectin; insecticide; emal1 protein.
XX
OS Streptomyces rimosus.
XX
PN WO200292801-A2.
XX
PD 21-NOV-2002.
XX
PF 15-MAY-2002; 2002WO-EP005363.
XX
PR 16-MAY-2001; 2001US-0291149P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
PI Buckel TG;
XX
DR WPI; 2003-140280/13.
DR N-PSDB; AAD53027.
XX
PT Novel polypeptide that exhibits enzymatic activity of P450 monooxygenase
PT and capable of regioselectively oxidizing alcohol of avermectin useful
PT for making emamectin from avermectin.
XX
PS Claim 17; Page 121-122; 157pp; English.
XX
CC The present invention relates to novel proteins that exhibit an enzymatic
CC activity of P450 monooxygenase and capable of regioselectively oxidising
CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
CC The invention also relates to ferrodoxins and ferrodoxin reductases that
CC are active with the P450 monooxygenases. Sequences of the invention are
CC useful for producing 4-keto-avermectin from avermectin, which is useful
CC for producing emamectin. Emamectin is useful as an insecticide. The
CC present sequence is Streptomyces rimosus emal1 protein
XX
SQ Sequence 475 AA;

Query Match 56.0%; Score 42; DB 6; Length 475;
Best Local Similarity 72.7%; Pred. No. 3.3e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 LPAVVGLSPGE 13
||: ||| |||
Db 49 LPSYVGLHPGE 59

RESULT 15
AAE34729
ID AAE34729 standard; protein; 475 AA.
XX
AC AAE34729;
XX
DT 14-MAY-2003 (first entry)

XX
DE Streptomyces albofaciens ema8 protein.
XX
KW P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;
KW emamectin; insecticide; ema8 protein.
XX
OS Streptomyces albofaciens.
XX
PN WO200292801-A2.
XX
PD 21-NOV-2002.
XX
PF 15-MAY-2002; 2002WO-EP005363.
XX
PR 16-MAY-2001; 2001US-0291149P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
PI Buckel TG;
XX
DR WPI; 2003-140280/13.
DR N-PSDB; AAD53024.
XX
PT Novel polypeptide that exhibits enzymatic activity of P450 monooxygenase
PT and capable of regioselectively oxidizing alcohol of avermectin useful
PT for making emamectin from avermectin.
XX
PS Claim 17; Page 116-117; 157pp; English.
XX
CC The present invention relates to novel proteins that exhibit an enzymatic
CC activity of P450 monooxygenase and capable of regioselectively oxidising
CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
CC The invention also relates to ferrodoxins and ferrodoxin reductases that
CC are active with the P450 monooxygenases. Sequences of the invention are
CC useful for producing 4-keto-avermectin from avermectin, which is useful
CC for producing emamectin. Emamectin is useful as an insecticide. The
CC present sequence is Streptomyces albofaciens ema8 protein
XX
SQ Sequence 475 AA;

Query Match 56.0%; Score 42; DB 6; Length 475;
Best Local Similarity 72.7%; Pred. No. 3.3e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy          3 LPAVVGLSPGE 13
              ||: ||| |||
Db         49 LPSYVGLHPGE 59
```

Search completed: May 7, 2004, 12:33:48
Job time : 43.85 secs